Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome

AUTHOR(S):

Bowman, John L.; Kohchi, Takayuki; Yamato, Katsuyuki T.; Jenkins, Jerry; Shu, Shengqiang; Ishizaki, Kimitsune; Yamaoka, Shohei; ... Watanabe, Yuichiro; Yazaki, Kazufumi; Yotsui, Izumi

CITATION:

Bowman, John L. ...[et al]. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell 2017, 171(2): 287-304

ISSUE DATE:

2017-10-05

URL:

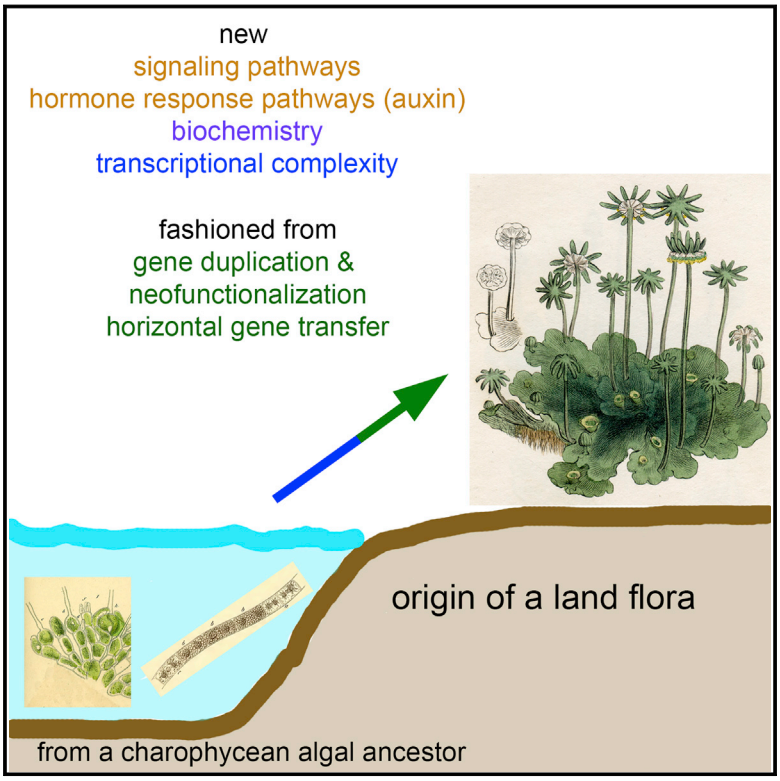
http://hdl.handle.net/2433/227477

Article

Cell

Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome

Graphical Abstract



Highlights

- First genome sequence of a liverwort, a basal lineage of land plants
- *Marchantia* genome content retains aspects of the genome of the ancestral land plant
- Evolution of haploid sex chromosomes in a haploid dominant dioecious plant
- F-box-mediated hormone signaling pathways evolved with origin of land plants

Authors

John L. Bowman, Takayuki Kohchi, Katsuyuki T. Yamato, ..., Izumi Yotsui, Sabine Zachgo, Jeremy Schmutz

Correspondence

john.bowman@monash.edu (J.L.B.), tkohchi@lif.kyoto-u.ac.jp (T.K.), kyamato@waka.kindai.ac.jp (K.T.Y.)

In Brief

The genome of the liverwort *Marchantia polymorpha* sheds light on the evolution of land plants.

Bowman et al., 2017, Cell *171*, 287–304
October 5, 2017
https://doi.org/10.1016/j.cell.2017.09.030

CellPress

Article

Cell

Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome

John L. Bowman,[1,49,*] Takayuki Kohchi,[2,*] Katsuyuki T. Yamato,[3,*] Jerry Jenkins,[4,5] Shengqiang Shu,[4] Kimitsune Ishizaki,[6] Shohei Yamaoka,[2] Ryuichi Nishihama,[2] Yasukazu Nakamura,[7] Frédéric Berger,[8] Catherine Adam,[4] Shiori Sugamata Aki,[9] Felix Althoff,[10] Takashi Araki,[2] Mario A. Arteaga-Vazquez,[11] Sureshkumar Balasubrmanian,[1] Kerrie Barry,[4] Diane Bauer,[4] Christian R. Boehm,[12] Liam Briginshaw,[1] Juan Caballero-Perez,[13] Bruno Catarino,[14] Feng Chen,[15] Shota Chiyoda,[2] Mansi Chovatia,[4] Kevin M. Davies,[16] Mihails Delmans,[12] Taku Demura,[9] Tom Dierschke,[1,10] Liam Dolan,[14] Ana E. Dorantes-Acosta,[11] D. Magnus Eklund,[1,17] Stevie N. Florent,[1] Eduardo Flores-Sandoval,[1] Asao Fujiyama,[7] Hideya Fukuzawa,[2] Bence Galik,[18] Daniel Grimanelli,[19] Jane Grimwood,[4,5] Ueli Grossniklaus,[20]

(Author list continued on next page)

[1]School of Biological Sciences, Monash University, Melbourne VIC 3800, Australia
[2]Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan
[3]Faculty of Biology-Oriented Science and Technology, Kindai University, 930 Nishimitani, Kinokawa, Wakayama 649-6493, Japan
[4]Department of Energy Joint Genome Institute, Walnut Creek, CA, USA
[5]HudsonAlpha Institute of Biotechnology, Huntsville, AL, USA
[6]Graduate School of Science, Kobe University, Kobe 657-8501, Japan
[7]National Institute of Genetics, Research Organization of Information and Systems, Yata, Mishima 411-8540, Japan
[8]Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[9]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara 630-0192, Japan
[10]Botany Department, University of Osnabrück, Barbarastr. 11, D-49076 Osnabrück, Germany
[11]Universidad Veracruzana, INBIOTECA - Instituto de Biotecnología y Ecología Aplicada, Av. de las Culturas Veracruzanas No.101, Colonia Emiliano Zapata, 91090, Xalapa, Veracruz, México
[12]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, United Kingdom
[13]National Laboratory of Genomics for Biodiversity, CINVESTAV-IPN, Km 9.6 Lib. Norte Carr. Irapuato-León, 36821, Irapuato, Guanajuato, México
[14]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK
[15]Department of Plant Sciences, University of Tennessee, Knoxville, TN, USA
[16]The New Zealand Institute for Plant & Food Research Limited, Private Bag 11-600, Palmerston North, New Zealand
[17]Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden
[18]Bioinformatics & Scientific Computing, Vienna Biocenter Core Facilities (VBCF), Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[19]Institut de Recherche pour le Développement (IRD), UMR232, Université de Montpellier, Montpellier 34394, France
[20]Department of Plant and Microbial Biology and Zurich-Basel Plant Science Center, University of Zurich, 8008 Zürich, Switzerland

(Affiliations continued on next page)

SUMMARY

The evolution of land flora transformed the terrestrial environment. Land plants evolved from an ancestral charophycean alga from which they inherited developmental, biochemical, and cell biological attributes. Additional biochemical and physiological adaptations to land, and a life cycle with an alternation between multicellular haploid and diploid generations that facilitated efficient dispersal of desiccation tolerant spores, evolved in the ancestral land plant. We analyzed the genome of the liverwort *Marchantia polymorpha*, a member of a basal land plant lineage. Relative to charophycean algae, land plant genomes are characterized by genes encoding novel biochemical pathways, new phytohormone signaling pathways (notably auxin), expanded repertoires of signaling pathways, and increased diversity in some transcription factor families. Compared with other sequenced land plants, *M. polymorpha* exhibits low genetic redundancy in most regulatory pathways, with this portion of its genome resembling that predicted for the ancestral land plant.

INTRODUCTION

Land plants evolved 450 mya from an ancestral charophycean alga from which they inherited numerous developmental, biochemical, and cell biological features (Figure 1; see Delwiche and Cooper [2015] for review). Characteristics of extant charophytes indicate that at least some terrestrial physiological adaptations (desiccation and UV radiation tolerance) evolved prior to land plants, and several land plant gene families originated in a charophycean ancestor (Hori et al., 2014; Ju et al., 2015). In

Cell

Takahiro Hamada,[21] Jim Haseloff,[12] Alexander J. Hetherington,[14] Asuka Higo,[2] Yuki Hirakawa,[1,22] Hope N. Hundley,[4] Yoko Ikeda,[23] Keisuke Inoue,[2] Shin-ichiro Inoue,[24] Sakiko Ishida,[2] Qidong Jia,[15] Mitsuru Kakita,[24] Takehiko Kanazawa,[25,26] Yosuke Kawai,[27] Tomokazu Kawashima,[8,28] Megan Kennedy,[4] Keita Kinose,[2] Toshinori Kinoshita,[22,24] Yuji Kohara,[7] Eri Koide,[2] Kenji Komatsu,[29] Sarah Kopischke,[10] Minoru Kubo,[9] Junko Kyozuka,[30] Ulf Lagercrantz,[17] Shih-Shun Lin,[31] Erika Lindquist,[4] Anna M. Lipzen,[4] Chia-Wei Lu,[31] Efraín De Luna,[32] Robert A. Martienssen,[33] Naoki Minamino,[25,26] Masaharu Mizutani,[34] Miya Mizutani,[2] Nobuyoshi Mochizuki,[35] Isabel Monte,[36] Rebecca Mosher,[37] Hideki Nagasaki,[7,38] Hirofumi Nakagami,[39,40] Satoshi Naramoto,[30] Kazuhiko Nishitani,[41] Misato Ohtani,[9] Takashi Okamoto,[42] Masaki Okumura,[24] Jeremy Phillips,[4] Bernardo Pollak,[12] Anke Reinders,[43] Moritz Rövekamp,[20] Ryosuke Sano,[9] Shinichiro Sawa,[44] Marc W. Schmid,[20] Makoto Shirakawa,[2] Roberto Solano,[36] Alexander Spunde,[4] Noriyuki Suetsugu,[2] Sumio Sugano,[45] Akifumi Sugiyama,[46] Rui Sun,[2] Yutaka Suzuki,[45] Mizuki Takenaka,[47] Daisuke Takezawa,[48] Hirokazu Tomogane,[2] Masayuki Tsuzuki,[21] Takashi Ueda,[25] Masaaki Umeda,[9] John M. Ward,[43] Yuichiro Watanabe,[21] Kazufumi Yazaki,[46] Ryusuke Yokoyama,[41] Yoshihiro Yoshitake,[2] Izumi Yotsui,[39] Sabine Zachgo,[10] and Jeremy Schmutz[4,5]

[21]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902 Japan
[22]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan; Department of Life Science, Faculty of Science, Gakushuin University, Tokyo 171-8588, Japan
[23]Institute of Plant Science and Resources, Okayama University, Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan
[24]Division of Biological Science, Graduate School of Science, Nagoya University, Chikusa, Nagoya 464-8602, Japan
[25]National Institute for Basic Biology, 38 Nishigounaka, Myodaiji, Okazaki 444-8585, Japan
[26]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[27]Department of Integrative Genomics, Tohoku Medical Bank Organization, Tohoku University, Aoba, Sendai 980-8573, Japan
[28]Department of Plant and Soil Sciences, University of Kentucky, 321 Plant Science Building, 1405 Veterans Dr., Lexington, KY 40546, USA
[29]Department of Bioproduction Technology, Junior College of Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan
[30]Graduate School of Life Sciences, Tohoku University, Sendai 980-8577, Japan
[31]Institute of Biotechnology, National Taiwan University, Taipei, Taiwan
[32]Instituto de Ecología, AC., Red de Biodiversidad y Sistemática, Xalapa, Veracruz, 91000, México
[33]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY, USA
[34]Graduate School of Agricultural Science, Kobe University, Rokkodai, Nada, Kobe 657-8501, Japan
[35]Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
[36]Department Genética Molecular de Plantas, Centro Nacional de Biotecnologia-CSIC, Universidad Autónoma de Madrid 28049 Madrid. Spain
[37]The School of Plant Sciences, The University of Arizona, Tuscon, AZ, USA
[38]Department of Technology Development, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan
[39]RIKEN Center for Sustainable Resource Science, Yokohama 230-0045, Japan
[40]Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany
[41]Laboratory of Plant Cell Wall Biology, Graduate School of Life Sciences, Tohoku University, Aoba, Sendai 980-8578, Japan
[42]Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan
[43]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, USA
[44]Graduate school of Science and Technology, Kumamoto University, Kurokami 2-39-1, Kumamoto 860-8555, Japan
[45]Department of Computational Biology and Medical Sciences, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562 Japan
[46]Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji 611-0011, Japan
[47]Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany
[48]Graduate School of Science and Engineering and Institute for Environmental Science and Technology, Saitama University, Saitama 338–8570, Japan
[49]Lead Contact
*Correspondence: john.bowman@monash.edu (J.L.B.), tkohchi@lif.kyoto-u.ac.jp (T.K.), kyamato@waka.kindai.ac.jp (K.T.Y.)
https://doi.org/10.1016/j.cell.2017.09.030

contrast, a defining land plant feature, a multicellular diploid sporophyte generation, was instrumental in land plants' ascent to terrestrial dominance, as it allowed optimized dispersal of desiccation tolerant spores (Bower, 1908).

The three early diverging extant land plant lineages (liverworts, mosses, hornworts; "bryophytes") lack vascular tissues and true roots but collectively possess all key innovations of land plant evolution: a multicellular diploid sporophyte, a gametophytic shoot apical meristem (SAM) with an apical cell producing 3-dimensional tissues, a sporophytic SAM, and cell fate specializations providing morphological and physiological terrestrial adaptations. Bryophyte phylogenetic relationships remain enigmatic, with nearly every possible topology proposed (Nishiyama et al., 2004; Qiu et al., 2006; Wickett et al., 2014). However, fossils assigned to liverworts (Ordovician-Silurian) predate those described as either mosses (Carboniferous) or hornworts (Cretaceous), suggesting that the first plants colonizing terrestrial habitats possessed attributes of liverworts (Edwards et al., 1995; Oostendorp, 1987; Wellman et al., 2003). Thus, liverworts may retain a larger suite of ancestral characters than other extant lineages.

Liverworts are monophyletic, with the dominant gametophyte body a dorsi-ventral thallus or leaf axis (Campbell, 1918). The sporophyte is reduced, compared with other bryophytes, and nutritionally relies on the gametophyte. Liverworts have

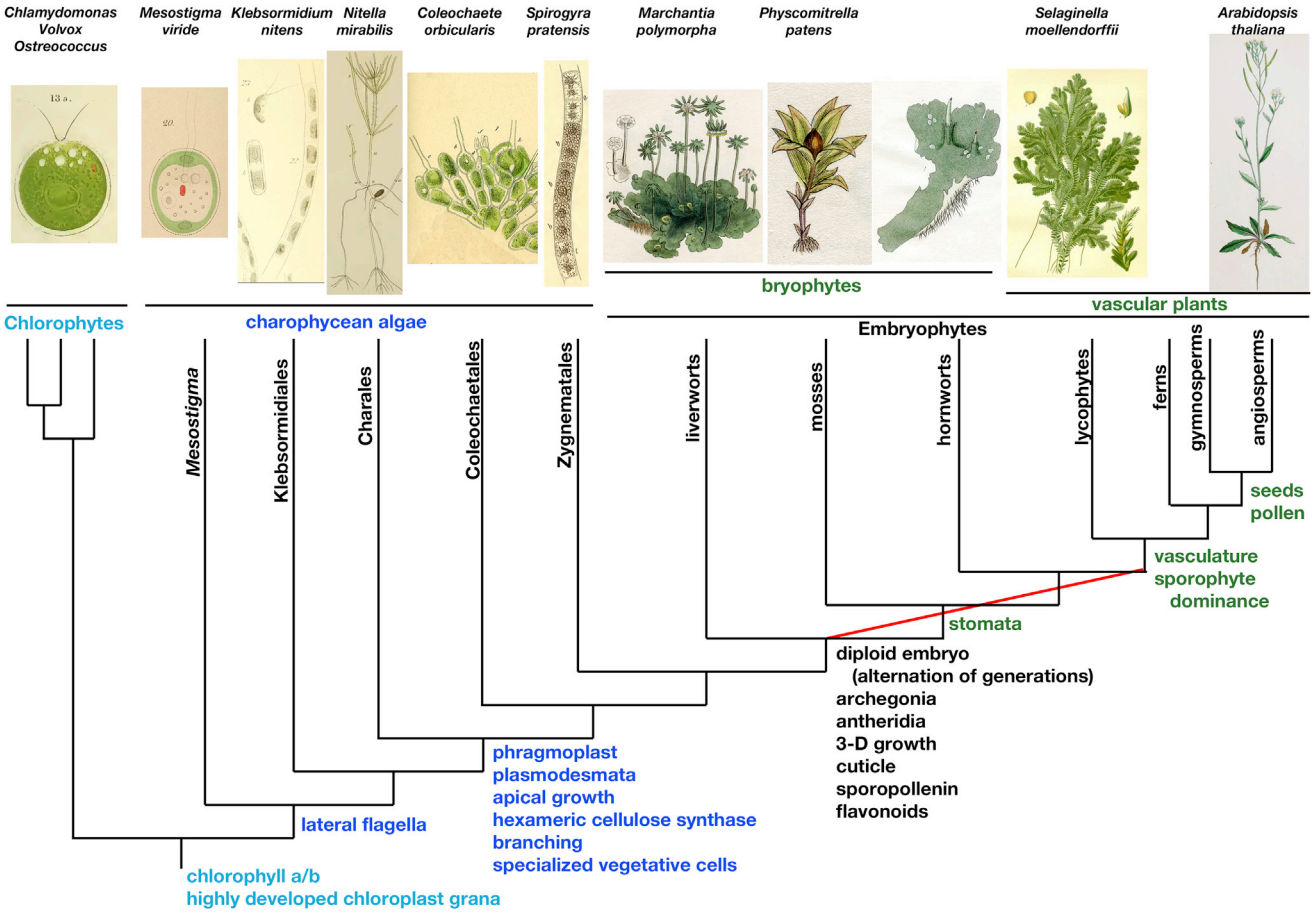


Figure 1. Phylogenetic Context of *Marchantia polymorpha*

Monophyletic land plants are nested within a charophycean algal grade, with phylogenetic relationships of bryophyte lineages unresolved (red line). Major evolutionary innovations within Viridiplantae are indicated.

membrane-bound oil bodies and ventral unicellular rhizoids and lack stomata. Liverworts exhibit a low rate of chromosomal evolution, with no evidence of ancient polyploidy (Berrie, 1960), and molecular evolution within the Marchantiopsida is slow compared with other liverwort lineages (Villarreal A et al., 2016). Dioecious liverworts possess sex chromosomes, with their presence possibly ancestral for the group.

Marchantia polymorpha is a member of the Marchantiopsida, a clade characterized by a complex thalloid gametophyte. Marchantialean fossils date to the Permo-Triassic period (Oostendorp, 1987), with the complex thallus a possible adaptation to arid conditions of the time (Wheeler, 2000). Due to ease of growth and genetic manipulation in the laboratory (Ishizaki et al., 2016) and an extensive historical literature (Bowman, 2016), we chose *M. polymorpha* as a representative liverwort, and we present an analysis of its genome.

RESULTS

Structural Genomics and Annotation

The nuclear and organellar genomes of *M. polymorpha* subspecies ruderalis (Bischler-Causse, 1993; Bowman et al., 2016a; Shimamura, 2016) were sequenced using a whole-genome shotgun sequencing strategy. We sequenced a single clonal female derived from backcross 4 between a male Tak-1 line, whose Y chromosome was previously sequenced (Yamato et al., 2007), and a female Tak-2 line, whose X chromosome was introgressed into a largely Tak-1 autosomal background. Nuclear genome v 3.1 assembly consists of 2,957 scaffolds (4,454 contigs) covering 225.8 Mb (see STAR Methods). The *M. polymorpha* plastid (120,304 bp) and mitochondrial (186,196 bp) genome sequences differ from previously published sequences (Oda et al., 1992; Ohyama et al., 1986), which are derived from *M. paleacea* rather than *M. polymorpha* (plastid, GenBank LC035012.1; [Kisiel et al., 2011]). Organellar genome organization is identical, and previous evolutionary and functional conclusions are not compromised.

Gene Numbers

Annotation revealed 19,138 nuclear encoded protein-coding genes, with 5,385 alternative protein-coding transcripts, (> 90%) having EST support (Table S1). To facilitate annotation, we obtained transcriptomes (see Table S2, STAR Methods, and Mendeley https://doi.org/10.17632/zb7hwyj3hp.1) representing various tissues of axenically grown plants, and from two "wild"

samples in which gene expression was detected for some genes without evidence of expression in axenically grown plants. With the arbitrary level ≥ 10-fold higher cut-off, sporophytic and antheridiophore (male reproductive) tissues possess more specialized transcriptomes than either sporeling or archegoniophore (female reproductive) tissues.

Highly expressed genes display a strong preference for synonymous cytosine in codon position 3, perhaps due to translational selection in such genes, as observed across a spectrum of plants (Wang and Roossinck, 2006). Average UTR and intron lengths of *M. polymorpha* genes are larger than in other land plants. 5′ UTRs contain a higher frequency of ATG(s) potentially encoding uORFs compared to other land plants examined (Table S1), suggesting that the translational machinery somehow recognizes the genuine *M. polymorpha* start codons.

The nuclear genome encodes 769 tRNA genes (51 pseudogenes) and 301 snRNA genes. Of 265 *M. polymorpha MIR* genes (Lin et al., 2016; Tsuzuki et al., 2016), 264 miRNA precursors (pre-miRNAs) were identified in diverse genomic contexts, including within protein-coding genes, with five present in tandem arrangements. ORFs (> 200 aa) were identified in 42 pre-miRNAs (Table S1). Expression patterns of pre-miRNAs negatively correlate with those of their targets, and *DCL1* orthologs—required for miRNA processing—were found in *M. polymorpha* and some charophytes.

Gene and Genome Evolution

6404 *M. polymorpha* genes were assigned to a eukaryotic orthologous group (KOG), 12,842 were assigned an OrthoMCL (with an overlap of 6,348 genes), and 419 additional genes assigned a PANTHER/Pfam category, leaving 5,821 genes lacking any annotation (Table S3). Comparative analysis across Viridiplantae revealed a larger increase of orthologous groups at the Streptophyta origin than at the land plant origin. Increased taxon sampling and analyses of shared algal and land plant orthologs may resolve bryophyte phylogeny. KOGs found in *M. polymorpha*, but not in other land plants, are often homologous with fungal genes or related to mobile elements, suggestive of horizontal gene transfer. The large number of orphan genes, lineage-specific, or evolutionarily recent genes (Tautz and Domazet-Lošo, 2011) may reflect the phylogenetic distance of *M. polymorpha* from other species with sequenced genomes.

Several gene families that are over-represented in *M. polymorpha* relative to other land plants encode transporters (e.g., phosphate and ammonium). Gene families missing in *M. polymorpha* include those required for successful arbuscular mycorrizal colonization, despite their presence in charophytes and closely related *Marchantia* species (Delaux et al., 2015). Related *Marchantia* form mycorrhizal associations, while *M. polymorpha* subsp. ruderalis does not (reviewed in Bowman et al. [2016a]). Increased transport capacity, rather than a reliance on mycorrhizal associations, could be a genomic adaptation for *M. polymorpha* being a weedy colonizer of barren disturbed habitats.

The ancestral chromosome number of all extant liverwort lineages is nine (Berrie, 1960; Heitz, 1927), implying an absence of ancient whole-genome duplications, and analyses confirm a lack of ancient polyploidy in *M. polymorpha* (Table S4). That a large majority of regulatory genes occur as single paralogs (see below) provides further support for this hypothesis and contributes to levels of paralogous genes in *M. polymorpha* (41.3%) being at the lower end of values observed in other land plants (45%–84%; Panchy et al. [2016]). However, local tandemly arrayed genes (TAGs) are not uncommon, with arrays of 2–11 adjacent paralogs accounting for 1,125 genes. The percentage of TAGs (5.9%) is at the lower end of the range observed in flowering plants (4.6%–26%; Panchy et al. [2016]; Rizzon et al. [2006]), but higher than in *Physcomitrella patens* (1%; Rensing et al. [2008]). In *M. polymorpha*, 75% of TAGs are encoded on the same strand, in contrast to *P. patens*, where a reduction of such TAGs was attributed to loss via homologous recombination (Rensing et al., 2008).

Repetitive DNA

Repetitive elements represent 22% of the *M. polymorpha* autosomal genome (Table S5), a value lower than that of *P. patens* (48%; Rensing et al. [2008]) but above that of the hornwort *Anthoceros agrestis* (6.98%; Szövényi [2016]). Similar to angiosperms, long terminal repeat (LTR) retroelements, including 264 full-length LTR retrotransposons, represent the largest fraction of repetitive elements (9.7%). X- and Y-specific repeat elements were previously reported (Yamato et al., 2007), and no new additional sex-specific elements were identified.

Sex Chromosomes

M. polymorpha possesses sex chromosomes, with an X chromosomal "feminizer" locus and multiple Y chromosomal loci required for sperm motility (see Bowman [2016] for review). Previous annotation of 6.0 Mb of the Y chromosome revealed 64 genes (Yamato et al., 2007). We identified 9 X chromosome scaffolds representing 4.37 Mb and annotated 74 X chromosome genes and 105 Y chromosome genes (see STAR Methods). X chromosome gene density (1 gene/57.5 kb) is similar to that of the Y (1 gene/56.7 kb)—about 5-fold lower than the autosomes (1 gene/11.3 kb). Of 74 X chromosome genes, 20 have their closest homologs on the Y chromosome and can be considered alleles (Figure 2). These genes are expressed vegetatively and conserved across land plants, Streptophyta or Viridiplantae, and thus represent relics of the ancestral autosome from which the sex chromosomes evolved (Table S6). Little evidence of synteny exists between X and Y scaffolds, suggesting an absence of recombination in these regions. Further, as noted previously (Yamato et al., 2007), synonymous substitutions between X and Y alleles are mutationally saturated, reflecting an ancient sex chromosome origin. Divergence between X and Y alleles is comparable to that between orthologs across extant liverwort, with phylogenetic analysis indicating an origin of sex chromosomes prior to Marchantiopsida diversification (Table S6), consistent with the idea that the ancestral liverwort may have possessed sex chromosomes (Allen, 1917; Berrie, 1960).

X or Y chromosome specific genes, when expressed, tend to be expressed in a reproductive-organ-specific manner (Table S6), conforming to predictions of sex chromosome evolution (Bull, 1978). However, many X-specific loci are genes or gene fragments with closely related autosomal paralogs, suggesting that they recently immigrated into the X chromosome (Table S6); a similar pattern has been reported in brown algae (Lipinska et al., 2017). Thus, only a handful of functionally X-specific genes exist, and among these, half have detectable expression only in

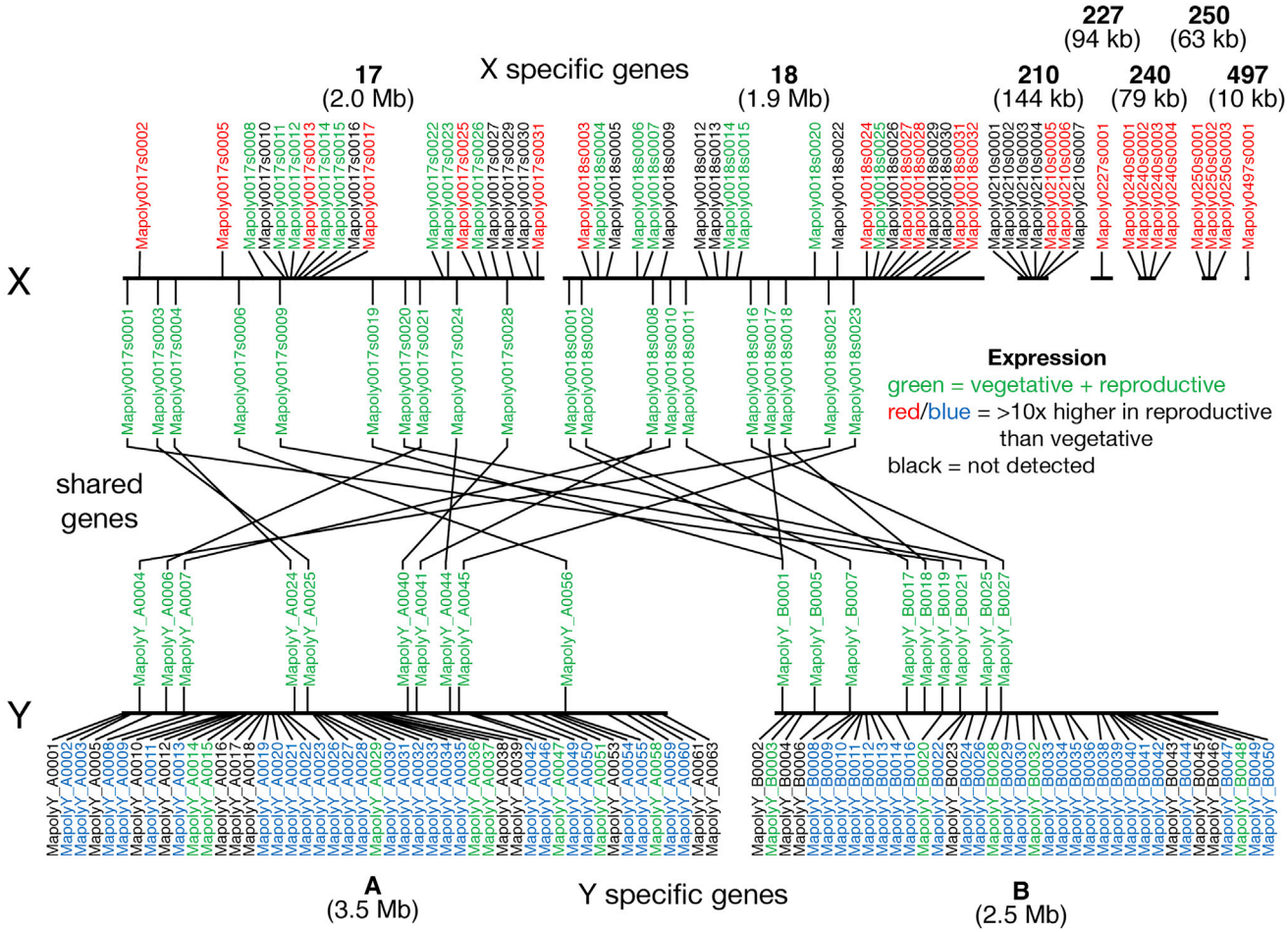


Figure 2. *M. polymorpha* Sex Chromosomes

Horizontal bars and vertical letters represent genomic sequences and genes, respectively. Shared genes are connected according to their phylogenetic relationships.

the sporophyte, leaving few obvious candidates for the feminizer locus. Y-specific genes exhibit a similar pattern, but include several conspicuous "motility" candidates.

Insights into Land Plant Evolution

In the remaining sections, we highlight two facets of land plant evolution gleaned via comparisons of the *M. polymorpha* genome content with genomes and transcriptomes of charophytes and other land plants. First, we detail genomic features that distinguish land plants from their algal relatives and describe their origins as precisely as possible. Origins are inferred in a manner analogous to the fossil record, i.e., they represent the latest possible dates unless orthologous algal sequences define origins precisely. Second, we note, where possible, the predicted ancestral land plant genome composition inferred from phylogenetic analyses.

Transcriptional Regulation

The *M. polymorpha* genome contains 394 (387 autosomal, 4 with X and Y alleles, and 3 X-specific) genes classified into 47 transcription factor (TF) families (Figure 3, Table S7). These TF families are present in other land plants, and no families present in other eukaryotes to the exclusion of Viridiplantae were found (Weirauch and Hughes, 2011). TFs in *M. polymorpha* account for 2.1% of protein coding genes, a lower percentage than in other land plants but higher than in algae, supporting the observation that TF numbers increase with organismal complexity (Catarino et al., 2016; Lang et al., 2010; Lehti-Shiu et al., 2017). An estimate of the TF content in the ancestral land plant was inferred via phylogenetic analyses of 18 TF families (Table S7, Figure 4, and Mendeley https://doi.org/10.17632/zb7hwyj3hp.1). The TF content of *M. polymorpha* resembles that predicted for the ancestral land plant, with a few exceptional lineage-specific gene expansions (trihelix, ASL/LBD, 3R-MYB).

TF diversity increased at the base of the Streptophyta prior to the divergence of *Klebsormidium* and the evolution of multicellularity (Figure 3, Table S7). Only one TF family evolved concomitantly with land plants (GeBP), and few families evolved within land plants (YABBY, VOZ, ULT). Thus, the origin of new TF families, per se, was not critical for land plant evolution. However,

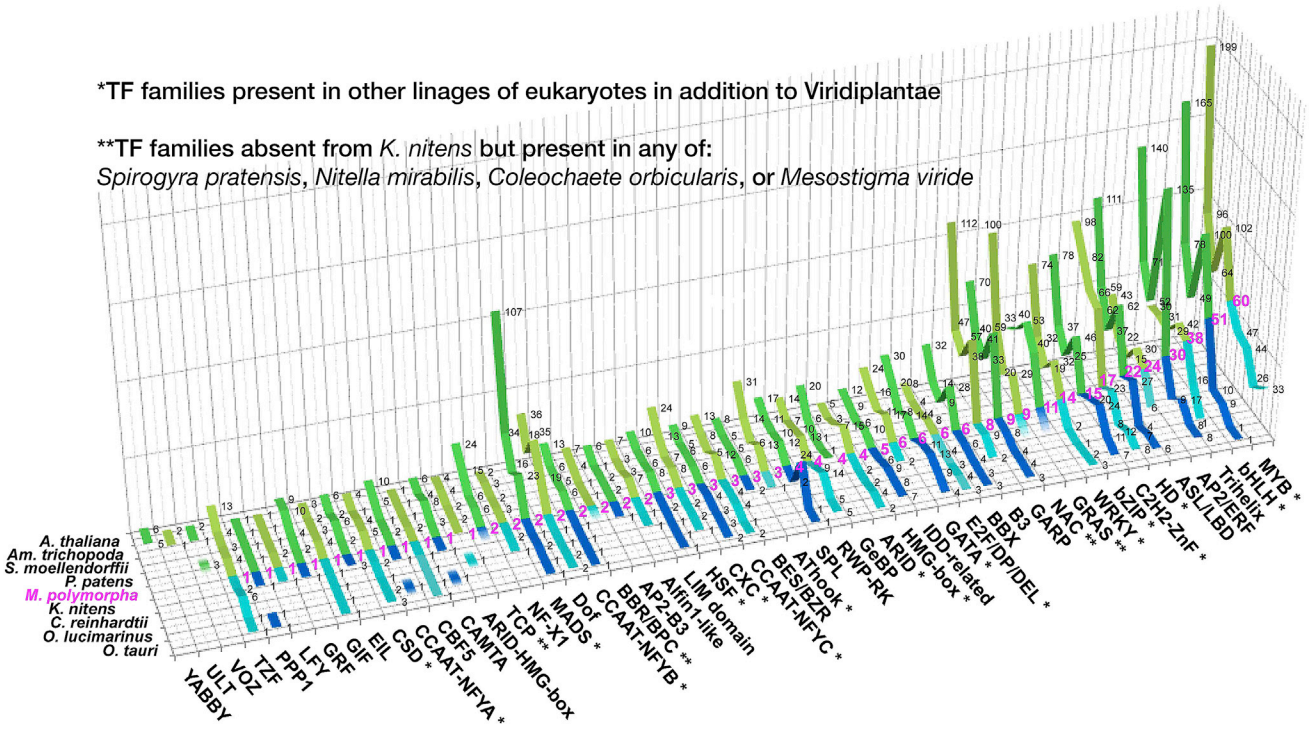


Figure 3. Numbers of Transcription Factor Genes in Land Plants and Algae

Transcription factor paralog numbers are listed for species (left) with sequenced genomes, with the green portion of the ribbon denoting presence in land plants and the blue presence in algae.

increased diversity of many TF families, such as bHLH, NAC, GRAS, AP2/ERF, ASL/LBD, and WRKY, suggests they may have been instrumental for terrestrialization. For example, 14 bHLH subfamilies were found exclusively in land plants (Figure S1B). Some subfamilies direct development of analogous cell types (angiosperm root hairs and bryophyte rhizoids) that are involved in nutrient and water uptake in land plants (Catarino et al., 2016). Other bHLH subfamilies function in differentiation of vascular plant tissue or cell types not found in liverworts (e.g., vasculature, stomata), indicating co-option of preexisting regulatory modules during land plant diversification.

Phylogenetic analysis of 18 TF families revealed four distinct recurring evolutionary patterns exemplified by clades of the MYB TF family (Figure 4), namely the following: (1) clades predating land plants; (2) clades evolving in the ancestral land plant, wherein a single *M. polymorpha* gene is orthologous to many other land plant paralogs; (3) clades for which there are no angiosperm orthologs of *M. polymorpha* genes; and (4) clades exhibiting a lineage specific expansion in *M. polymorpha.* Pattern 2 is exemplified by clades 1, 4, and 5 (Figure 4), whose members function in regulating secondary metabolism. Pattern 4 is exemplified by 3R-MYBs.

Pattern 3 is notable in the RWP-RK (Koi et al., 2016; Rövekamp et al., 2016) and TALE class homeodomain families, which are phylogenetically more diverse in *M. polymorpha* (and other basal land plant lineages) than in angiosperms. A striking example of the opposite pattern is the MADS-box gene family, with type II MADS-box genes diversifying within land plants (Figure 3) and type I genes lacking in *M. polymorpha* despite being found in other land plant lineages.

Sporophyte Expression of Transcription Factors

A defining feature of land plants is a multicellular sporophyte (Hofmeister, 1862). Evolution from an ancestral single-celled zygote to a multicellular sporophyte was likely accomplished via changes in existing, and establishment of new, gene regulatory networks. In two related mosses, a largely overlapping set of TFs is predominantly sporophytically expressed (O'Donoghue et al., 2013; Ortiz-Ramírez et al., 2016; Szövényi et al., 2011). In *M. polymorpha*, 41 TFs are predominantly expressed during sporophyte development, but only 10 are potential orthologs of similarly expressed moss genes (Table S7). Of these, two are TALE-HD genes that are known to regulate the alternation of generations (Bowman et al., 2016b). The minimal overlap possibly reflects the significant divergence in sporophyte anatomy and morphology between liverworts and mosses (Campbell, 1918).

Chromatin

M. polymorpha possesses homologs of most chromatin-related genes identified in other eukaryotes (Table S7). A proportion of genes functioning in chromatin structure and modification are highly expressed in either male reproductive (33%) or sporophytic (19%) tissues. Preferential or unique sporophytic expression is frequently observed in families with multiple paralogs (13 families out of 18), suggesting that changes in chromatin occur during the gametophyte to sporophyte transition, as noted in *P. patens* (Mosquna et al., 2009; Okano et al., 2009).

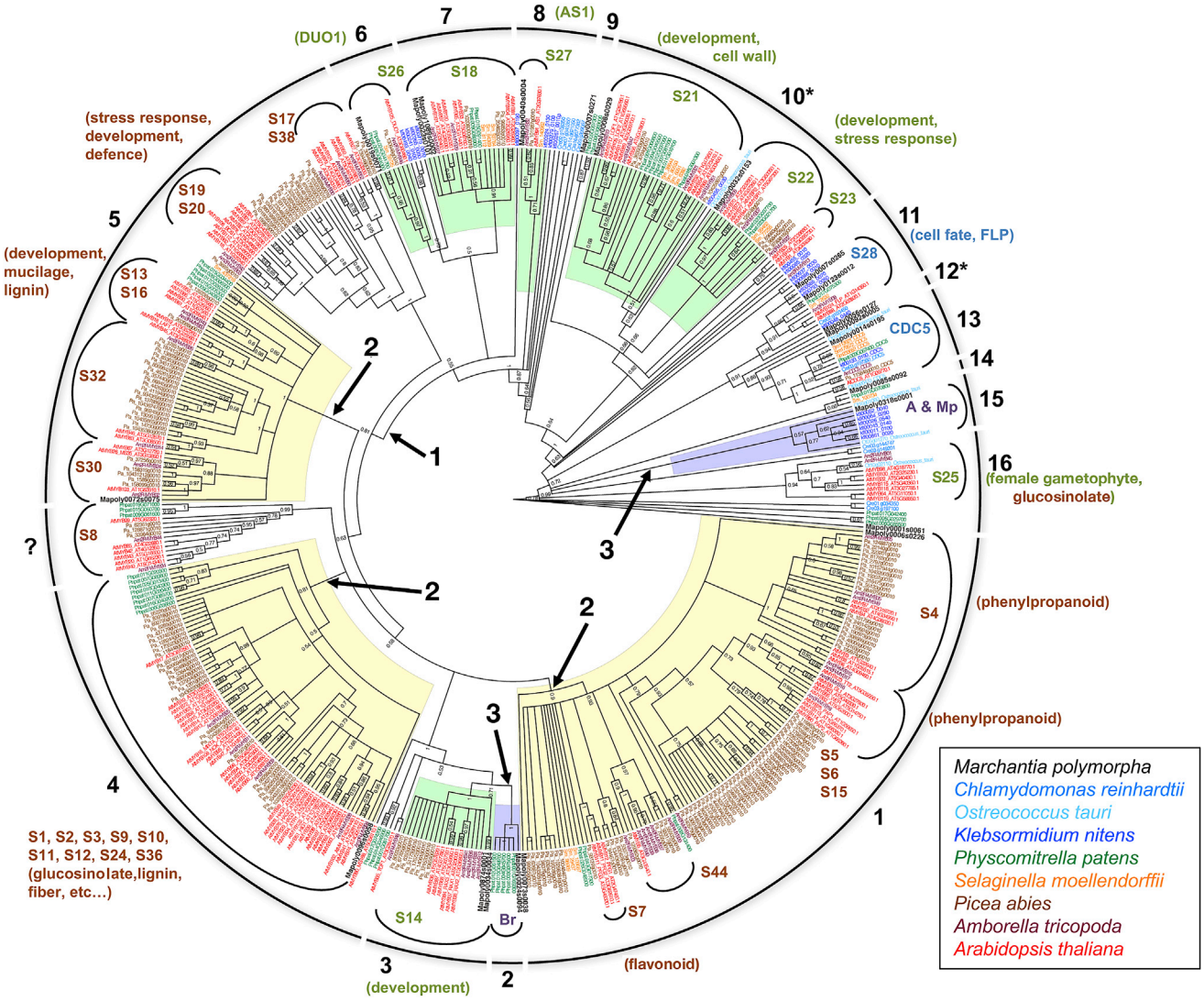


Figure 4. Phylogenetic Relationships of R2R3-MYBs

Central black numbers refer to modes of evolution described. Peripheral numbers indicate clades predicted in the ancestral land plant, with a potential loss in *M. polymorpha* (?) and poorly resolved clades with potentially > 1 ancestral land plant gene (*). Green clades largely control development processes, while yellow clades concern secondary metabolism. S numbers refer to previous classification of R2R3-MYBs (Du et al., 2015; Dubos et al., 2010); however, angiosperm-based classifications are often inadequate in a broader phylogenetic context.

Chromatin-related genes could also account for specific sporophytic features, such as mitosis in diploid cells or a possible requirement for imprinting or dosage compensation. Similarly, regulation of chromosome compaction might explain specific expression patterns during male gametogenesis.

DNA methylation is a heritable eukaryotic epigenetic modification that plays roles in repetitive element silencing and regulation of some protein coding genes. The *M. polymorpha* vegetative gametophyte showed DNA methylation enrichment and 23–24 nt siRNA clusters at genomic loci with repetitive elements, matching the pattern in vascular plants, but not over gene bodies (Takuno et al., 2016), with the exception of genes on sex chromosomes, where methylation might spread from nearby repetitive elements (Table S6). *M. polymorpha* possesses a suite of methyltransferases similar to other land plants (Table S7), except an absence of CMT3, which may be related to lack of gene body methylation (Takuno et al., 2016).

RNA Biology

Genes encoding SR family proteins mediating mRNA splicing are found in single copy in *M. polymorpha* (Table S7). Pentatricopeptide repeat (PPR) proteins direct RNA processing steps in organelles, with land-plant-specific PLS-class PPR proteins mediating RNA editing (Fujii and Small, 2011). The single *M. polymorpha* PLS gene lacks domains characterizing other RNA-editing PLS proteins (Table S7), consistent with a loss of RNA editing in Marchantiopsida organelles (Rüdinger et al., 2008) and implying a different function for the *M. polymorpha* PLS protein.

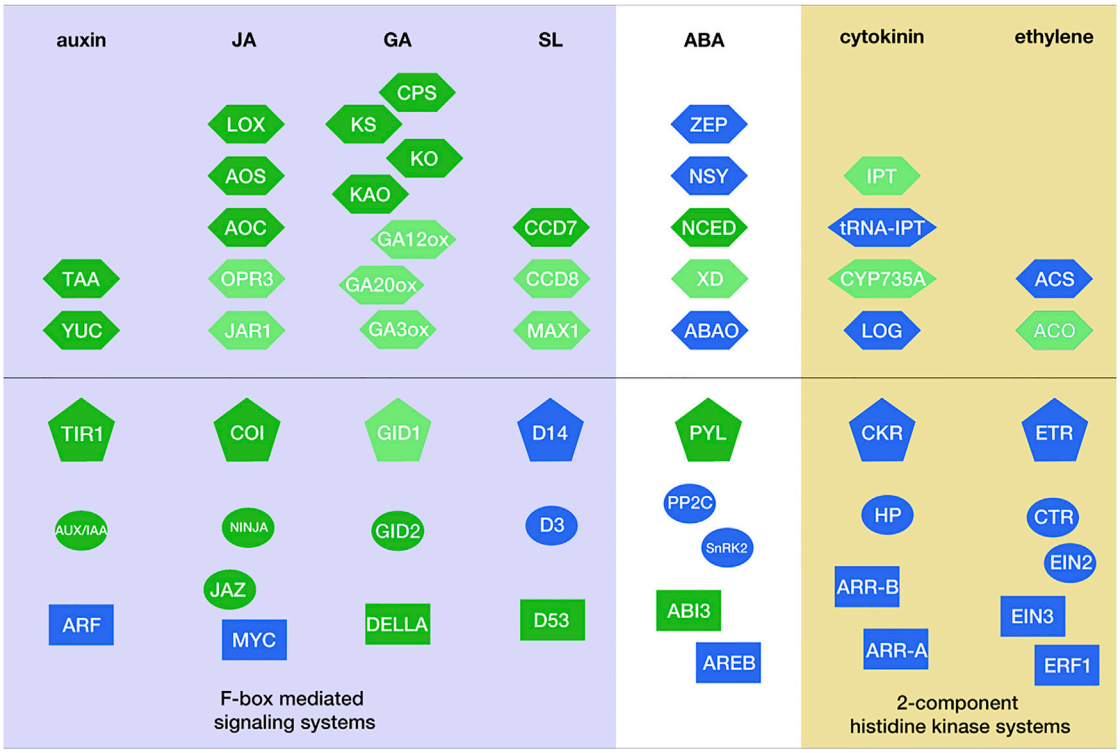


Figure 5. Origin of Phytohormone Biosynthesis and Signaling Pathways

Land plant signaling pathways comprise very ancient two-component system hormone pathways involved in cytokinin and ethylene signal transduction, which evolved in an ancestral alga, and F-box mediated pathways involving the plant hormones auxin, jasmonate, and strigolactones that evolved in land plants from pre-existing components. Even when signaling pathways are conserved, biosynthetic pathways may vary between algae and land plants or even within land plants. Biosynthesis enzymes are shown as hexagons, receptors as pentagons, signal transduction components as ovals, and transcription factors as rectangles; blue, present in algae; dark green, origin with land plants; light green, origin within land plants.

Diversification of Signaling Pathways

One hallmark of multicellular life is a plethora of signaling pathways by which cells communicate, influencing cell specification, differentiation, and physiology. In most cases, the *M. polymorpha* genome encodes a minimal, but complete, set of land-plant-signaling molecules.

F-box-mediated Hormone Pathways

Auxin modulates growth, differentiation, and gravitational and light-tropic responses. The conserved land plant auxin biosynthetic pathway via indole-3-pyruvic acid (Eklund et al., 2015) consists of one autosomal Mp*TAA* gene and five YUCCA paralogs, with one expressed gametophytically, in *M. polymorpha*. Chlorophytes and charophytes possess a gene that is orthologous to a land plant clade containing both TAA and its paralog TAR (Table S8), but no charophycean orthologs of land plant YUCCA genes are recognizable (Yue et al., 2012), suggesting that this auxin biosynthetic pathway evolved in land plants.

The *M. polymorpha* auxin transcriptional response machinery consists of single orthologs of AUX/IAA, TOPLESS (TPL), three classes of ARF genes, and the TIR1 receptor (Flores-Sandoval et al., 2015; Kato et al., 2015). At least two ARF classes exist in charophytes, with the association of B3 DNA binding and PB1 protein interaction domains arising early in streptophyte evolution (Figure S1). AUX/IAA genes were derived from an ancestral ARF via B3 domain loss and gain of domains (I, II) that interact with the TPL corepressor and the TIR1 receptor plus auxin. Charophyte proteins with PB1 domains that are most similar to land plant AUX/IAA proteins lack domains I and II (Figure S1). Thus, while charophytes have orthologs of land plant AUX/IAA genes, they lack specific domains required for canonical auxin transcriptional responses. In *M. polymoprha*, two additional genes (Mp*AXI1*, Mp*AXI2*) encode PB1 domains but lack the B3 domain, possibly providing additional ARF regulation. Charophytes have an F-box gene that is orthologous to both TIR1 (the auxin receptor) and COI (the jasmonate receptor) (Table S8). However, these algal F-box orthologs do not possess conserved key residues for interaction with corresponding land plant ligands (JA-Ile/JAZ [Sheard et al., 2010] and auxin and AUX/IAA [Tan et al., 2007]; Figure S2), suggesting that they are neither auxin nor JA receptors. In spite of previous assumptions (Wang et al., 2015), it appears that both auxin and JA perception machineries evolved in the ancestral land plant from an ancestral algal F-box via gene duplication and neofunctionalization (Figures 5 and S3).

Polar auxin transport (PAT) occurs in liverwort gametophytes (LaRue and Narayanaswami, 1957; Maravolo, 1976). PAT depends on PIN and ABCB efflux and AUX1 influx facilitators. Both PIN and ABCB orthologs exist in *Klebsormidium*, raising

the posibility of PAT in charophytes (Hori et al., 2014). *M. polymorpha* has 5 PIN genes and several ABCB auxin efflux facilitator orthologs. A single AUX1 ortholog is present in *M. polymorpha*, but no similar charophycean algal sequence was identified. AGC VIII group kinases and NPH3/RPT2-like proteins that regulate ABCB and PIN localization and activity arose prior to the colonization of land but diversified in the ancestral land plant (Suetsugu et al., 2016). Thus, land plant auxin efflux facilitators and transporters consist of both ancient and more recently evolved components, implying that PAT predated land plants and increased in complexity during land plant evolution.

Plants respond to biotic and abiotic stresses via the jasmonate (JA) signaling pathway. Similar to other bryophytes, *M. polymorpha* can synthesize the JA-precursor OPDA (12-oxo-phytodienoic acid) but lacks OPR3 (OPDA reductase) that produces the vascular plant hormone JA-Ile (Stumpe et al., 2010; Yamamoto et al., 2015). Despite this, single orthologs of the COI1 co-receptor, JAZ repressor, MYC transcription factor, and NINJA adaptor to the TPL co-repressor are found in *M. polymorpha* (Table S8). No clear JAZ or NINJA orthologs exist in either Chlorophytes or charophytes. Thus the JA-perception and repression machinery (COI1-JAZ, NINJA) arose in the land plant ancestor, with pre-existing MYC TFs recruited to modulate activity by a newly acquired JA- or OPDA-related ligand.

Strigolactones (SLs) are secreted signals inhibiting shoot branching and promoting arbuscular mycorrhizal symbiosis. Bryophyte SL biosynthesis pathways differ from those of vascular plants (Delaux et al., 2012; Proust et al., 2011). A minimal set of SL signaling components exist in *M. polymorpha*, with only some components present in charophytes, suggesting that the land plant SL pathway was constructed from preexisting and newly evolved components (Table S8). Similarly, homologs for some, but not all, vascular plant GA-synthesizing enzymes were identifed in *M. polymorpha*, consistent with the absence of canonical GA compounds in bryophytes. Orthologs of GA-signaling pathway components, except GID1, are present in *M. polymorpha*; thus, compared to *P. patens*, the *M. polymorpha* GA signaling module is closer to that in vascular plants, but lack of GID1 suggests stepwise acquisition during land plant evolution.

Abscisic Acid

Abscisic acid (ABA) regulates dormancy and stress acclimation under water-limited environments in land plants (Sakata et al., 2014). *M. polymorpha* produces endogenous ABA, and gemmae undergo growth inhibition with desiccation tolerance upon exogenous ABA treatment (Akter et al., 2014; Li et al., 1994; Tougane et al., 2010). Single orthologs of ABA biosynthetic enzymes, except XD, ABA catabolism, and ABCG transporters for ABA import and export are present in *M. polymorpha* (Table S8). A PTR-type transporter was not identified. Cellular ABA response is mediated by an intracellular receptor (PYL), wherein the ABA-PYL complex activates SNF1-related protein kinase2 (SnRK2) through inhibition of group A protein phosphatase 2C (PP2C). Mp*PYL1* is a functional receptor, as it complements *Arabidopsis pyr1pyl1pyl2pyl4* mutants (Figure S4). *M. polymorpha* has a diversity of PYL paralogs that is not observed in other land plants, and subfunctionalization of MpPYL paralogs is evident in sporophyte-specific expression patterns. PP2C (Tougane et al., 2010) and SnRK2 exist throughout Viridiplantae, but PYL receptors evolved more recently. Thus, acquisition of the PYL receptor was crucial for recruitment of ABA as a phytohormone in an ancestor of land plants. Downstream TF ABI3, which plays a key role in desiccation tolerance in bryophytes, evolved with land plants, while AREB arose in streptophytes.

Two Component System Hormone Pathways

In contrast to F-box-mediated hormone pathways, two-component system cytokinin and ethylene signal transduction pathways are present in charophytes, but not Chlorophytes, suggesting an origin early in charophyte evolution (Table S8, Ju et al. [2015]). As in other non-seed plants (Banks et al., 2011; Rensing et al., 2008), genes encoding a distinct ethylene-forming enzyme (ACO) that is characteristic of seed plants are not present in *M. polymorpha*.

Receptor Kinase Signaling Pathways

Cells perceive extracellular molecules via transmembrane receptors, and a large numerical increase of RLK/Pelle kinases in land plants compared to Chlorophytes has been documented (Lehti-Shiu and Shiu, 2012). The paraphyletic leucine-rich repeat receptor kinase family (LRR-RLK) is the largest subclass of land plant RLK/Pelle kinases (Shiu and Bleecker, 2001). 107 *M. polymorpha* genes possess both LRR and kinase domains, with 14 out of 15 subclasses characterized in *Arabidopsis* present (Table S9), with some notable absences, e.g., *BRI1* and *PSKR1*. Charophytes possess members in only a few subclasses, indicating diversification of LRR-RLK genes in the ancestral land plant. At least 15 peptide families act as signals in *Arabidopsis* (Matsubayashi, 2014). Members of the CLE, IDA, EPFL, and RALF families are present in *M. polymorpha*, but we could not detect other families (Table S9). Receptors for identified peptides (CLV1, HAESA, ERECTA, and FERONIA) are found in *M. polymorpha*, but not in charophytes, indicating that these peptide-signaling pathways evolved in an ancestral land plant. Similarly, most RLK/Pelle kinase subclasses (Lehti-Shiu and Shiu, 2012) found throughout land plants are present in *M. polymorpha* but missing in charophytes (Table S9), implying diversification of both developmental and defense signaling systems with the advent of terrestrialization.

In contrast to RLK/Pelle kinases, most subfamilies of MAPK(K)(K) genes are present in charophytes or Chlorophytes, indicating that most extant MAPK cascade diversity evolved prior to land plants, as did cell cycle machinery (Table S9). Likewise, all subfamilies of land plant PPP and PP2C phosphatases are already present in charophytes or Chlorophytes. MAPK and phosphatase gene familes in the *M. polymorpha* genome are encoded minimal paralogs compared with other land plants and are similar to predictions for the ancestral land plant. In contrast, histidine kinase (HK) diversity arose early in charophyte evolution, and portions were lost prior to and during land plant evolution. Among other signaling pathways, only one class of F-box originated with land plants, and ancestral land plant PEPB diversity was higher than in extant angiosperms.

New Features Adaptive to Life on Land

The transition to land from a previously aquatic or semi-aquatic habitat entailed adaptation to a host of environmental

challenges, both abiotic and biotic, requiring new or modified biochemistry.

Prevention against Photooxidative Stress

Plants evolved distinct photoreceptor families for adaptation to ambient light conditions and efficient photosynthesis, with all angiosperm photoreceptor classes predating land plant evolution. *M. polymorpha* has single orthologs of photoreceptors and core light signaling components resembling those predicted for the ancestral land plant and, in addition, a couple liverwort-specific LOV domain proteins (Table S10). Phenylpropanoids, in particular flavonoids, act as UV "sunscreens" in land plants, and their production is upregulated by UV-B via the UVR8 photoreceptor. Liverworts produce a variety of small phenylpropanoid metabolites, including anti-fungal bibenzyls, UV-absorbing flavone glycosides, yellow aurone glycosides, and cell-wall-located red pigmentation, presumably the anthocyanidins riccionidin A and B (Kunz et al., 1993). Genes encoding core enzymes of the phenylpropanoid and shikimate pathways and UVR8 signaling are present in *M. polymorpha*, primarily in single or low copy number, with exceptions being PAL and CHS found in tandem arrays (Table S10). There are no convincing reports of flavonoids from algae, and we find no evidence of phenylpropanoid biosynthetic genes in charophytes. At least some of these enzymes (e.g., PAL) may be derived via horizontal gene transfer from soil microbes to the ancestral land plant (Emiliani et al., 2009).

Control of Hydration

Cell wall rigidity and imperviousness enhances efficient water conduction and retention. Typical land-plant cell-wall biochemistry exists in charophytes, with wall composition and mechanical properties proposed to reflect an early adaptation to terrestrialization (Harholt et al., 2016; Mikkelsen et al., 2014). Annotation of *M. polymorpha* genes encoding cell-wall-related enzymes supports this view, with most glycoside transferase and hydrolase families present in charophytes (Table S10). Diversification of carbohydrate esterases, polysaccaride lyases, pectin methyl transferases, and origin of xyloglucan endotransglucosylase and hydrolases (XTH) coincide with land plant emergence. Lack of charophyte XTHs is consistent with reports they lack xyloglucan. MpXTH and expansin diversity suggests that remodeling of the xyloglucan-cellulose network may be more important than pectin in regulating *M. polymorpha* cell wall mechanical properties compared with seed plants (Hongo et al., 2012).

Water distribution involves, in part, plasmodesmata, plasma membrane-lined intercellular connections that evolved within charophytes or in the ancestral land plant (Brunkard and Zambryski, 2017). Most known angiosperm plasmodesmatal protein homologs are restricted to vascular plants or have broad distributions, including species lacking plasmodesmata; thus, their origins either predate or postdate plasmodesmata evolution; only two exceptions appear to be land plant-specific (Table S10). Plasmodesmata formation is associated with the unique type of cytokinesis mediated by cell plate formation that evolved in the derived streptophytes (Pickett-Heaps, 1969) and is correlated with diversification of membrane trafficking machinery in charophytes (Sanderfoot, 2007). Evolution of a streptophyte-specific clade of SYP1 genes, including SYP11/KNOLLE, a cytokinesis-specific syntaxin, may have been instrumental in cell division mechanism evolution. Unlike other land plants, liverworts also retain ancestral centrosome-like polar organizers and thus represent a transitional state between ancestral centripedal cleavage and derived centrifugal division mechanisms (Brown and Lemmon, 1990; Buschmann et al., 2016; Farmer, 1895).

The polymer lignin increases rigidity and imperviousness of secondary cell walls—a key innovation of water-conducting systems and structural support. While *M. polymorpha* possesses all putative lignin biosynthesis genes except ferulate 5-hydroxylase (Table S10), similar to *P. patens* (Xu et al., 2009), bona fide lignin polymers have not been found. Thus, these enzymes may function in a "pre-lignin" pathway as suggested for *P. patens* (Renault et al., 2017). Differentiation of water-conducting cells in other land plants is controlled by VNS subfamily NAC TFs that upregulate cell wall and programmed cell death (PCD) genes (Xu et al., 2014). While *M. polymorpha* has no internal water-conducting cells, it possesses the core set of VNS-downstream genes that could function in cells with secondary thickenings or undergo PCD (elaters and pegged rhizoids). Charophytes possess homologs of VNS-downstream enzymes but lack the NAC and MYB orthologs (Figure 4, Table S10), implying that the VNS regulatory network was established in the ancestral land plant and later co-opted to direct conducting cell differentiation in derived lineages.

M. polymorpha dorsal thallus surfaces are hydrophobic, attributable to a cuticle preventing water loss and providing protection from insects and UV radiation (Schönherr and Ziegler, 1975). Land plant cuticles consist of a cutin polymer forming an extracellular structural matrix with interspersed cuticular waxes. Orthologs of genes regulating cuticle biosynthesis and deposition are present in *M. polymorpha* (Table S10). In contrast, homologs for monomer biosynthesis (CYP86A) were not identified, consistent with cutin monomers in Marchantiopsida differing from those of other land plants (Caldicott and Eglinton, 1976). A similar picture emerges for homologs of wax biosynthesis. Despite potentially analogous surface layers in some charophytes (Cook and Graham, 1998), no closely related cutin biosynthesis or transporter homologs were identified in charophytes.

The polymer sporopollenin is a major component of spore exine, protecting spores from desiccation. During sporopollenin biosynthesis, hydroxylated alkylpyrones are generated by acyl-CoA synthetase, anther-specific chalcone synthase-like, and tetraketide α-pyrone reductase (Daku et al., 2016), with *M. polymorpha* possessing candidates for all enzymes. Plants synthesize a diverse array of additional specialized metabolites arising, in part, from the functions of cytochrome P450 monooxygenases (CYPs), 2-oxoglutarate-dependent dioxygenases (2OGDs), and UDP-dependent glycosyltransferases (UGTs). In contrast to regulatory genes, the majority of *M. polymorpha* 148 CYP, 41 UGT, and 38 2OGD genes represent novel families, indicating substantial lineage-specific diversification of specialized metabolism. In addition to water, channels also facilitate regulated sensing and uptake of exongenous molecules and their distribution throughout the plant. Most angiosperm membrane-associated channels and transporters have *M. polymorpha* homologs (Table S10), with most diversity arising

in algal ancestors, with exceptions being NPF transporters and a subclade of ABCG transporters that diversified in the ancestral land plant.

Defense

The signaling molecule salicylic acid (SA) regulates defense responses against pathogens (Dempsey et al., 2011). Homologs of all key components in SA biosynthesis and signaling were identified in *M. polymorpha* (Table S10). Both the PAL-dependent SA biosynthesis pathway and the SA transcriptional response mediator, *NPR1*, are found only in land plants. Expansion of the downstream *PR1* (pathogenesis-related) gene family in land plants may have followed *NPR1* gene acquisition.

Plants sense pathogens via two types of receptors. Plasma membrane located pattern recognition receptors, FLS2 and EFR, belong to the LRR-RLK subfamily XII and require SERK (LRR-RLK subfamily II) co-receptors (Chinchilla et al., 2007; Gómez-Gómez and Boller, 2000; Roux et al., 2011; Zipfel et al., 2006). *M. polymorpha* does not have *FLS2* or *EFR* orthologs (Table S10), but other LRR-RLK subfamily XII genes could have roles in sensing bacterial elicitors. In contrast, *SERK* genes are conserved among land plants. Intracellular receptors, containing nucleotide-binding site (NBS) and LRR domains (NBS-LRR), directly or indirectly recognize pathogen virulence molecules. A molecular chaperone complex required for NBS-LRR protein activity predates land plants (Table S10). *NBS-LRR* genes evolve rapidly with large clusters of genes in *Arabidopsis* (536) and *P. patens* (165) (Sarris et al., 2016). *M. polymorpha* possesses only 34 genes, some of which are paired, suggesting a similar genomic architecture as in angiosperms. The absence of known NBS-LRR downstream signaling genes in non-seed plants suggests that this pathway diversified later.

Oil bodies, unique to liverworts, accumulate terpenoids and serve a role in deterring pathogens and herbivores (Suire et al., 2000; Tanaka et al., 2016). *M. polymorpha* possesses 7 typical plant terpene synthase and 32 microbial terpene synthase-like (MTPSL) genes, including 10 partial genes. Several *M. polymorpha* MTPSLs function as sesquiterpene or monoterpene synthases in *in vitro* assays (Kumar et al., 2016). The ancestral MTPSL genes were acquired through horizontal gene transfer from bacteria and fungi (Jia et al., 2016), providing liverworts with a chemical defense to deter herbivores and pathogens from their inception (Labandeira et al., 2014).

Horizontal Gene Transfer

Given the apparent horizontal gene transfer (HGT) of MTPSL, we searched for families with a similar phylogenetic distribution—in *M. polymorpha* and fungi, to the exclusion of all other eukaryotes. Including MTPSLs, we identified 10 families (42 genes) most closely related to fungal genes and 5 families (31 genes) most closely related to bacterial genes as candidates for HGT from microbes to liverworts (Table S11), a number that is similar to that reported for *P. patens* (Yue et al., 2012). Sequence similarity to either Ascomycota or Basidiomycota implies multiple transfers not acquired from mycorrhizal fungi associates, which are typically Glomeromycota in complex thalloid liverworts (Field et al., 2015). While the majority of genes suggested by Yue et al. (2012) to be acquired by the ancestral land plant are present in charophycean algae (Table S11), given the potential HGT frequency, further analyses may reveal additional events in the ancestral land plant as it was exposed to new microbes during terrestrialization.

DISCUSSION

Genome Composition

Our data are consistent with an ancient origin of dimorphic sex chromosomes within the liverwort lineage. In organisms with a dominant diploid generation, the unique heterogametic sex chromosome undergoes degeneration due to accumulation of detrimental mutations via a process first outlined by Muller (Muller, 1914). In contrast, the evolutionary fate of sex chromosomes in organisms with a dominant haploid generation is fundamentally different, with predictions being the following: (1) the sex chromosomes should have similar characteristics, with degeneration similar for both; (2) degeneration should be limited, with retention of genes required gametophytically and loss of sporophytic genes; and (3) changes in size should be additions of heterochromatin (Bull, 1978). If the ancestral autosome from which *M. polymorpha* sex chromosomes evolved resembled extant autosomes, an aspect of theory not supported is limited degeneration, as sex chromosome gene density implies gene loss with concomitant repetitive elements acquisition.

One unique feature of the *M. polymorpha* genome relative to other sequenced land plant genomes is lack of redundancy in most regulatory genes. Regardless of the phylogenetic position of liverworts, this implies that the last common ancestor of extant land plants likely possessed a regulatory genome similar to that of *M. polymorpha*. While *M. polymorpha* has a paucity of duplicated regulatory factors, other biosynthetic, metabolic, and structural genes do not exhibit this pattern. How can this be explained? Whole-genome duplications (WGDs), from *Paramecium* to angiosperms, result in retention of an over-representation of TFs and regulatory molecules (Edger and Pires, 2009; McGrath et al., 2014; Papp et al., 2003). In contrast, due to dosage sensitivity, small-scale duplications of TF genes are selected against because their products often act in complexes, while small-scale duplications of biosynthesis, metabolic, and structural genes are not (Birchler et al., 2001; Hanada et al., 2008; Maere et al., 2005). Since there is little evidence for ancient WGDs within liverworts (Berrie, 1960; Heitz, 1927), the pattern of retained *M. polymorpha* paralogs may be a reflection of the lack of ancient WGDs. The question then becomes, "why were there no ancient WGDs in liverworts?" Mable lists several conditions that may either limit or promote polyploidy (Mable, 2004). Notably, most factors suggested to facilitate polyploid formation exist in liverworts: e.g., high rates of vegetative propagation, self-compatibility, and a measurable rate of dyad formation. Muller hypothesized that polyploidy was less common in animals than in plants due to the presence of strongly dimorphic sex chromosomes whose segregation during meiosis in tetraploids leads to non-viable chromosome constitutions (Muller, 1925). While the presence of sex chromosomes does not preclude polyploidy (Mable, 2004), we propose that the early evolution of dimorphic sex chromosomes contributed to genome stability in liverworts. The formation of meiotic dyads in the Marchantiopsida is not rare, but the lack of ancient polyploidy indicates selection against polyploid genotypes, perhaps due to inefficient pairing

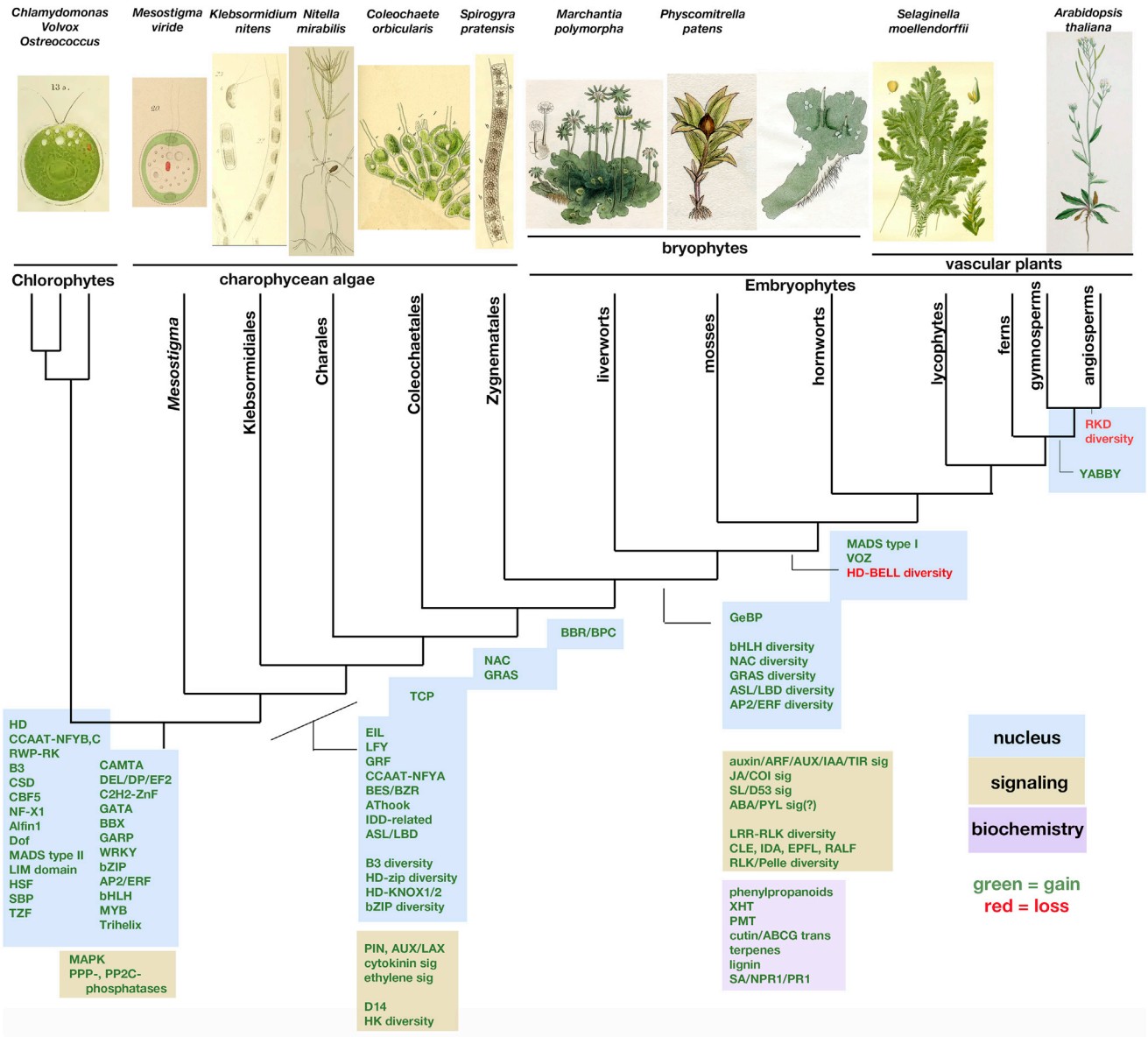


Figure 6. Evolution of Land Plant Attributes

The acquisition, or loss, of genes involved in transcription, signaling, and biochemistry during the evolution of land plants from an algal ancestor are shown, with positions reflecting the latest possible common ancestor.

of the sex chromosomes, as observed in induced liverwort polyploids (Allen, 1935; Haupt, 1932; Lorbeer, 1927). In cases where liverworts have recently, e.g., within extant genera, experienced polyploidy, the species almost invariably become monoecious (Berrie, 1964; Heitz, 1927). Consistent with our hypothesis, mosses, ferns, and angiosperms do not have ancient dimorphic sex chromosomes and readily become polyploid, while hornworts, which may possess sex chromosomes, are rarely polyploid (Berrie, 1960; Rink, 1935). In this scenario, the early evolution of sex chromosomes in the liverwort lineage resulted in a lack of WGDs, which in turn reduced proliferation of regulatory genes. Whether retention of a largely "ancestral regulatory genome" has restricted or canalized liverwort evolution is a matter for debate (Van de Peer et al., 2017); however, the diversity of liverwort gametophyte morphology compared to that of mosses and hornworts (Campbell, 1918) might argue against the premise.

Land Plant Evolution

Previous genomic comparisons have hinted at genetic innovations that define land plants (Floyd and Bowman, 2007; Hori et al., 2014; Ju et al., 2015; Nishiyama et al., 2003; Rensing et al., 2008). Using genomic and transcriptomic resources spanning charophyte and land plant diversity, we refined the

evolutionary origins of numerous gene families, with signaling pathways prominent among land plant innovations. Compared with those of their charophycean algal relatives, land plant genomes are characterized by genes encoding novel biochemical pathways (cutin, phenylpropanoids, terpenoids, lignin, SA), new F-box-mediated phytohormone signaling pathways (auxin, JA, SL), expanded repetoires of receptor-like kinases (RLK/Pelle) and their peptide ligands (CLE, IDA, EPFL, RALF), but not MAPK or phosphatase families, and increased diversity in a few TF families (Figure 6). In the simplest cases, gene duplication followed by neofunctionalization was sufficient to drive molecular diversification (TFs, cutin, RLK/Pelle). However, in other cases, neofunctionalization of proteins to assemble *de novo* molecular interactions between preexisting and newly evolved proteins was required, e.g., in the assembly of the auxin and JA pathways. As auxin influences a vast array of developmental mechanisms in land plants, the evolution of its transcriptional signaling pathway was critical to set the stage for the morphological diversification of land plants. Finally, interactions with other terrestrial organisms not only resulted in the establishment of symbiotic interactions (e.g., mycorrhizal fungi [Field et al., 2015]), but also access to genes acquired via horizontal transfer that provided novel biochemistry (terpenoids, phenylpropanoids, auxin biosynthesis). These innovations were critical for both developmental evolution of land plant architecture and adaption to new abiotic (desiccation, UV, temperature fluctuation) and biotic stressors typical of life on land.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and 11 tables and can be found with this article online at https://doi.org/10.1016/j.cell.2017.09.030.

An audio PaperClip is available at https://doi.org/10.1016/j.cell.2017.07.036#mmc12.

AUTHOR CONTRIBUTIONS

Conceptualization, J.L.B., and T. Kohchi; Methodology, J.L.B., T. Kohchi, K.T.Y., R.N., and J.S.; Software, K.T.Y.; Validation, J.L.B., T. Kohchi, K.T.Y., J.J., S. Shu, K. Ishizaki, S.Y., R.N., Y.N., F.B., and J.S.; Investigation, J.L.B., T. Kohchi, K.T.Y., J.J., S. Shu, K. Ishizaki, S.Y., R.N., Y.N., F.B., C.A., S.S.A., F.A., T.A., M.A.A.-V., S.B., K.B., D.B., C.R.B., L.B., J.C.-P., B.C., F.C., M.C., K.M.D., M.D., T. Demura, T. Dierschke, L.D., A.E.D.-A., D.M.E., S.N.F., E.F.-S., D.G., J.G., U.G., T.H., J.H., A.J.H., A.H., Y.H., H.N.H., Y.I., K. Inoue, S.-i.I., S.I., Q.J., M. Kakita, T. Kanazawa, Y. Kawai, T. Kawashima, M. Kennedy, K. Kinose, T. Kinoshita, E.K., K. Komatsu, S.K., M. Kubo, J.K., U.L., S.-S.L., E.L., A.M.L., C.-W.L., E.L., R.M., N. Minamino, Masaharu Mizutani, N. Mochizuki, I.M., R.M., H. Nakagami, S.N., K.N., M. Ohtani, T.O., M. Okumara, J.P., B.P., A.R., M.R., R. Sano, S. Sawa, M.W.S., M.S., R. Solano, A. Spunde, N.S., A. Sugiyama, R. Sun, M. Takenaka, D.T., H.T., M. Tsuzuki, T.U., M.U., J.M.W., Y.W., K.Y., R.Y., Y.Y., I.Y., S.Z., and J.S.; Resources, K. Ishizaki, S.C., A.F., H.F., B.G., S.I., Y. Kohara, Miya Mizutani, H. Nagasaki, S. Sugano, Y.S., and J.S.; Data Curation, J.L.B., K.T.Y., Y.N., and H. Nagasaki; Writing – Original Draft, J.L.B., T. Kohchi, K.T.Y., K. Ishizaki, S.Y., R.N., F.B., S.S.A., F.A., M.A.A.-V., C.R.B., F.C., K.M.D., L.D., U.G., T.H., Y.H., Y.I., S.-i.I., Q.J., T. Kanazawa, T. Kawashima, T. Kinoshita, K. Komatsu, J.K., S.-S.L., N. Minamino, Masaharo Mizutani, N. Mochizuki, I.M., H. Nakagami, S.N., K.N., M. Ohtani, A.R., R. Solano, N.S., A. Sugiyama, R. Sun, M. Takenaka, D.T., M. Tsuzuki, T.U., M.U., J.M.W., Y.W., K.Y., R.Y., Y.Y., and S.Z.; Writing – Review & Editing, J.L.B., T. Kohchi, K.T.Y., K. Ishizaki, S.Y., R.N., and F.B.; Visualization, J.L.B., T. Kohchi, K.T.Y., K. Ishizaki, S.Y., and R.N.; Supervision, J.L.B., T. Kohchi, K.T.Y., and F.B.; Project Administration, J.L.B., T. Kohchi, and K.T.Y.; Funding Acquisition, J.L.B., T. Kohchi, K.T.Y., K. Ishizaki, F.B., M.A.A.-V., C.R.B., K.M.D., A.E.D.-A., D.G., U.G., J.H., Y.I., E.L., and J.S.

ACKNOWLEDGMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Funding from ARC FF0561326, DP130100177, DP160100892 (J.L.B), MEXT KAKENHI Grant Numbers (15K21758, 16H06279, 25113001, 25113009 [T. Kohchi], 22112514, 24112715, 24510272 [K.T.Y.], 15H04391, 15H01233, 17H06472 [K. Ishizaki], FWF grant P28320-B21 (F.B.), SNF grant 310030B_160336 (U.G.), Gates Cambridge Trust (C.R.B.), Program to Disseminate Tenure Tracking System, MEXT, Japan (Y.I.), the Marsden Fund of New Zealand grant PAF1302 (K.D.), UC MEXUS Collaborative program grant 2011-UCMEXUS-19941-44-OAC7 (M.A.A.-V, X.C., and E.D.L.), Consejo Nacional de Ciencia y Tecnología (CONACYT) grant CB-158550 (M.A.A.-V.), COSEAMX1 JEAI EPIMAIZE, Universidad Veracruzana (M.A.A.-V. and D.G.), CONACYT grant CB-158561 (A.E.D.A.), Newton Fund grant RG79985 (M.A.A.-V. and J.H.), and Universidad Veracruzana - Cuerpo Académico CA-UVER-234, Marie Sklodowska-Curie action (#658900) (D.G.) is acknowledged. We thank Sandra K. Floyd for DNA

and RNA isolation; David Powell, Kathryn Hodgins, and Xuemei Chen for bioinformatics assistance; and Kanako Takeda, Manami Miyazaki, Takumi Kondoh, Yuiri Sowa, Yuma Hibi, Ryo Hirai, Motoyoshi Nakamura, Jun Inoue, Tatsuya Hasegawa, Mihoko Kobayashi, Kyoko Miyashita, and Taisuke Togawa for isolation of X-linked markers. We apologize to those researchers whose works we consulted but were unable to cite due to lack of space.

Received: December 11, 2016
Revised: April 21, 2017
Accepted: September 18, 2017
Published: October 5, 2017

REFERENCES

Akter, K., Kato, M., Sato, Y., Kaneko, Y., and Takezawa, D. (2014). Abscisic acid-induced rearrangement of intracellular structures associated with freezing and desiccation stress tolerance in the liverwort Marchantia polymorpha. J. Plant Physiol. *171*, 1334–1343.

Albert, V.A., Barbazuk, W.B., dePamphilis, C.W., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., et al.; Amborella Genome Project (2013). The Amborella genome and the evolution of flowering plants. Science *342*, 1241089.

Allen, C.E. (1917). A chromosome difference correlated with sex differences in sphærocarpos. Science *46*, 466–467.

Allen, C.E. (1935). The Occurrence of Polyploidy in Sphaerocarpos. Am. J. Bot. *22*, 664–680.

Altenhoff, A.M., Škunca, N., Glover, N., Train, C.M., Sueki, A., Piližota, I., Gori, K., Tomiczek, B., Müller, S., Redestig, H., et al. (2015). The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res. *43*, D240–D249.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science *332*, 960–963.

Berrie, G.K. (1960). The chromosome numbers of liverworts (Hepaticae and Anthocerotae). Trans. Brit. Bryol. Soc. *3*, 688–705.

Berrie, G.K. (1964). Experimental Studies on Polyploidy in Liverworts I. The Riccia fluitans Complex. Bryologist *67*, 146–152.

Birchler, J.A., Bhadra, U., Bhadra, M.P., and Auger, D.L. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev. Biol. *234*, 275–288.

Bischler-Causse, H. (1993). Marchantia L. The European and African taxa, *Volume 45* (Berlin, Stuttgart: J. Cramer).

Bower, F.O. (1908). Origin of a land flora: a theory based on the facts of alternation (London: MacMillan and Co.).

Bowman, J.L. (2016). A Brief History of Marchantia from Greece to Genomics. Plant Cell Physiol. *57*, 210–229.

Bowman, J.L., Araki, T., Arteaga-Vazquez, M.A., Berger, F., Dolan, L., Haseloff, J., Ishizaki, K., Kyozuka, J., Lin, S.S., Nagasaki, H., et al. (2016a). The Naming of Names: Guidelines for Gene Nomenclature in Marchantia. Plant Cell Physiol. *57*, 257–261.

Bowman, J.L., Sakakibara, K., Furumizu, C., and Dierschke, T. (2016b). Evolution in the cycles of life. Annu. Rev. Genet. *50*, 133–154.

Brown, R.C., and Lemmon, B.E. (1990). Polar organizers mark division axis prior to preprophase band formation in mitosis of the hepatic Reboulia hemisphaerica (Bryophyta). Protoplasma *156*, 74–81.

Brunkard, J.O., and Zambryski, P.C. (2017). Plasmodesmata enable multicellularity: new insights into their evolution, biogenesis, and functions in development and immunity. Curr. Opin. Plant Biol. *35*, 76–83.

Bull, J.J. (1978). Sex Chromosomes in Haploid Dioecy- A Unique Contrast to Muller's Theory for Diploid Dioecy. Am. Nat. *112*, 245–250.

Buschmann, H., Holtmannspötter, M., Borchers, A., O'Donoghue, M.T., and Zachgo, S. (2016). Microtubule dynamics of the centrosome-like polar organizers from the basal land plant Marchantia polymorpha. New Phytol. *209*, 999–1013.

Caldicott, A.B., and Eglinton, G. (1976). Gas Chromatographic-Mass Spectrometric Studies of Long-Chain Hydroxy-Acids 0.7. Cutin Acids from Bryophytes - Omega-1 Hydroxy Alkanoic Acid in 2 Liverwort Species. Phytochemistry *15*, 1139–1143.

Campbell, D.H. (1918). Mosses and Ferns, 2nd ed. (New York: The Macmillan Company).

Catarino, B., Hetherington, A.J., Emms, D.M., Kelly, S., and Dolan, L. (2016). The Stepwise Increase in the Number of Transcription Factor Families in the Precambrian Predated the Diversification of Plants On Land. Mol. Biol. Evol. *33*, 2815–2819.

Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. *34*, D363–D368.

Chevreux, B., Wetter, T., and Suhai, S. (1999). Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) *99*, 45–56.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature *448*, 497–500.

Cook, M.E., and Graham, L.E. (1998). Structural similarities between surface layers of selected charophycean algae and bryophytes and the cuticles of vascular plants. Int. J. Plant Sci. *159*, 780–787.

Cooper, E., and Delwiche, C. (2016). Green algal transcriptomes for phylogenetics and comparative genomics. figshare. https://figshare.com/articles/Green_algal_transcriptomes_for_phylogenetics_and_comparative_genomics/1604778.

Daku, R.M., Rabbi, F., Buttigieg, J., Coulson, I.M., Horne, D., Martens, G., Ashton, N.W., and Suh, D.Y. (2016). PpASCL, the Physcomitrella patens Anther-Specific Chalcone Synthase-Like Enzyme Implicated in Sporopollenin Biosynthesis, Is Needed for Integrity of the Moss Spore Wall and Spore Viability. PLoS ONE *11*, e0146817.

Delaux, P.M., Xie, X., Timme, R.E., Puech-Pages, V., Dunand, C., Lecompte, E., Delwiche, C.F., Yoneyama, K., Bécard, G., and Séjalon-Delmas, N. (2012). Origin of strigolactones in the green lineage. New Phytol. *195*, 857–871.

Delaux, P.M., Radhakrishnan, G.V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J.D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., et al. (2015). Algal ancestor of land plants was preadapted for symbiosis. Proc. Natl. Acad. Sci. USA *112*, 13390–13395.

Delwiche, C.F., and Cooper, E.D. (2015). The Evolutionary Origin of a Terrestrial Flora. Curr. Biol. *25*, R899–R910.

Dempsey, D.A., Vlot, A.C., Wildermuth, M.C., and Klessig, D.F. (2011). Salicylic Acid biosynthesis and metabolism. Arabidopsis Book *9*, e0156.

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc. Natl. Acad. Sci. USA *103*, 11647–11652.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Du, H., Liang, Z., Zhao, S., Nan, M.G., Tran, L.S.P., Lu, K., Huang, Y.B., and Li, J.N. (2015). The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. Sci. Rep. *5*, 11037.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends Plant Sci. *15*, 573–581.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Edger, P.P., and Pires, J.C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res. *17*, 699–717.

Edwards, D., Duckett, J.G., and Richardson, J.B. (1995). Hepatic Characters in the Earliest Land Plants. Nature *374*, 635–636.

Eklund, D.M., Ishizaki, K., Flores-Sandoval, E., Kikuchi, S., Takebayashi, Y., Tsukamoto, S., Hirakawa, Y., Nonomura, M., Kato, H., Kouno, M., et al. (2015). Auxin Produced by the Indole-3-Pyruvic Acid Pathway Regulates Development and Gemmae Dormancy in the Liverwort Marchantia polymorpha. Plant Cell *27*, 1650–1669.

El Baidouri, M., Kim, K.D., Abernathy, B., Arikit, S., Maumus, F., Panaud, O., Meyers, B.C., and Jackson, S.A. (2015). A new approach for annotation of transposable elements using small RNA mapping. Nucleic Acids Res. *43*, e84.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics *9*, 18.

Emiliani, G., Fondi, M., Fani, R., and Gribaldo, S. (2009). A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol. Direct *4*, 7.

Farmer, J.B. (1895). On Spore-Formation and Nuclear Division in the Hepaticae. Ann. Bot. (Lond.) *9*, 469–523.

Field, K.J., Rimington, W.R., Bidartondo, M.I., Allinson, K.E., Beerling, D.J., Cameron, D.D., Duckett, J.G., Leake, J.R., and Pressel, S. (2015). First evidence of mutualism between ancient plant lineages (Haplomitriopsida liverworts) and Mucoromycotina fungi and its response to simulated Palaeozoic changes in atmospheric CO2. New Phytol. *205*, 743–756.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Flores-Sandoval, E., Eklund, D.M., and Bowman, J.L. (2015). A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort Marchantia polymorpha. PLoS Genet. *11*, e1005207.

Flores-Sandoval, E., Dierschke, T., Fisher, T.J., and Bowman, J.L. (2016). Efficient and Inducible Use of Artificial MicroRNAs in Marchantia polymorpha. Plant Cell Physiol. *57*, 281–290.

Floyd, S.K., and Bowman, J.L. (2007). The ancestral developmental tool kit of land plants. Int. J. Plant Sci. *168*, 1–35.

Fujii, S., and Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol. *191*, 37–47.

Fujisawa, M., Hayashi, K., Nishio, T., Bando, T., Okada, S., Yamato, K.T., Fukuzawa, H., and Ohyama, K. (2001). Isolation of X and Y chromosome-specific DNA markers from a liverwort, Marchantia polymorpha, by representational difference analysis. Genetics *159*, 981–985.

Gómez-Gómez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol. Cell *5*, 1003–1011.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.

Haas, B.J., Delcher, A.L., Wortman, J.R., and Salzberg, S.L. (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics *20*, 3643–3646.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol. *148*, 993–1003.

Harholt, J., Moestrup, Ø., and Ulvskov, P. (2016). Why Plants Were Terrestrial from the Beginning. Trends Plant Sci. *21*, 96–101.

Haupt, G. (1932). Beiträge zur Zytologie der Gattung Marchantia I. Z. Indukt. Abstamm. Vererbungsl. *62*, 367–428.

Heitz, E. (1927). Über multiple und aberrante Chromosomenzahlen. Abhandlungen Naturwissenschatlicher Verein Hamburg *21*, 48–58.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics *32*, 767–769.

Hofmeister, W.F.B. (1862). On the germination, development, and fructification of the higher Cryptogamia, and on the fructification of the Coniferae (London: Ray Society).

Hongo, S., Sato, K., Yokoyama, R., and Nishitani, K. (2012). Demethylesterification of the primary wall by PECTIN METHYLESTERASE35 provides mechanical support to the Arabidopsis stem. Plant Cell *24*, 2624–2634.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. *5*, 3978.

Ishizaki, K., Nishihama, R., Yamato, K.T., and Kohchi, T. (2016). Molecular Genetic Tools and Techniques for Marchantia polymorpha Research. Plant Cell Physiol. *57*, 262–270.

Jaffe, D.B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J.P., Zody, M.C., and Lander, E.S. (2003). Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res. *13*, 91–96.

Jia, Q., Li, G., Köllner, T.G., Fu, J., Chen, X., Xiong, W., Crandall-Stotler, B.J., Bowman, J.L., Weston, D.J., Zhang, Y., et al. (2016). Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. Proc. Natl. Acad. Sci. USA *113*, 12328–12333.

Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nat Plants *1*, 14004.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J.L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS Genet. *11*, e1005084.

Kent, W.J. (2002). BLAT–the BLAST-like alignment tool. Genome Res. *12*, 656–664.

Kisiel, K., Miwa, H., and Odrzykoski, I.J. (2011). Taxonomic identification of chloroplast genome of Marchantia polymorpha using DNA barcode sequences. In Fourth International Barcode of Life Conference, pp. A43.

Koi, S., Hisanaga, T., Sato, K., Shimamura, M., Yamato, K.T., Ishizaki, K., Kohchi, T., and Nakajima, K. (2016). An Evolutionarily Conserved Plant RKD Factor Controls Germ Cell Differentiation. Curr. Biol. *26*, 1775–1781.

Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. *5*, R7.

Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics *23*, 1026–1028.

Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., Bell, S.A., Nybo, S.E., Kinison, S.E., Jiang, Z., et al. (2016). Molecular diversity of terpene synthases in the liverwort Marchantia polymorpha. Plant Cell *28*, 2632–2650.

Kunz, S., Burkhardt, G., and Becker, H. (1993). Riccionidins a and b, anthocyanidins from the cell walls of the liverwort Ricciocarpos natans. Phytochemistry *35*, 233–235.

Labandeira, C.C., Tremblay, S.L., Bartowski, K.E., and VanAller Hernick, L. (2014). Middle Devonian liverwort herbivory and antiherbivore defence. New Phytol. *202*, 247–258.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riaño-Pachón, D.M., Corrêa, L.G.G., Reski, R., Mueller-Roeber, B., and Rensing, S.A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol. Evol. *2*, 488–503.

LaRue, C.D., and Narayanaswami, S. (1957). Auxin Inhibition in the Liverwort Lunularia. New Phytol. *56*, 61–70.

Lehti-Shiu, M.D., and Shiu, S.H. (2012). Diversity, classification and function of the plant protein kinase superfamily. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 2619–2639.

Lehti-Shiu, M.D., Panchy, N., Wang, P., Uygun, S., and Shiu, S.-H. (2017). Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. Biochim. Biophys. Acta. *1860*, 3–20.

Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity (Edinb) *104*, 520–533.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, X., Wurtele, E.S., and Lamotte, C.E. (1994). Abscisic-Acid Is Present in Liverworts. Phytochemistry *37*, 625–627.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lin, P.C., Lu, C.W., Shen, B.N., Lee, G.Z., Bowman, J.L., Arteaga-Vazquez, M.A., Liu, L.Y., Hong, S.F., Lo, C.F., Su, G.M., et al. (2016). Identification of miRNAs and Their Targets in the Liverwort Marchantia polymorpha by Integrating RNA-Seq and Degradome Analyses. Plant Cell Physiol. *57*, 339–358.

Lipinska, A.P., Toda, N.R.T., Heesch, S., Peters, A.F., Cock, J.M., and Coelho, S.M. (2017). Multiple gene movements into and out of haploid sex chromosomes. Genome Biol. *18*, 104.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–D495.

Lorbeer, G. (1927). Untersuchungen über Reduktionsteilung und Geschlechtsbestimmung bei Lebermoosen. Z. Indukt. Abstamm. Vererbungsl. *44*, 1–109.

Lyons, E., Pedersen, B., Kane, J., and Freeling, M. (2008). The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. Trop. Plant Biol. *1*, 181–190.

Mable, B.K. (2004). 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biol. J. Linn. Soc. Lond. *82*, 453–466.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. USA *102*, 5454–5459.

Maravolo, N.C. (1976). Polarity and Localization of Auxin Movement in Hepatic, Marchantia polymorpha. Am. J. Bot. *63*, 526–531.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S.N., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). Cdd: Ncbi's Conserved Domain Database. Nucleic Acids Res. *43*, D222–D226.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. Annu. Rev. Plant Biol. *65*, 385–413.

McGrath, C.L., Gout, J.F., Johri, P., Doak, T.G., and Lynch, M. (2014). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome Res. *24*, 1665–1675.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., et al. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245–250.

Mikkelsen, M.D., Harholt, J., Ulvskov, P., Johansen, I.E., Fangel, J.U., Doblin, M.S., Bacic, A., and Willats, W.G.T. (2014). Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae. Ann. Bot. (Lond.) *114*, 1217–1236.

Mosquna, A., Katz, A., Decker, E.L., Rensing, S.A., Reski, R., and Ohad, N. (2009). Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development *136*, 2433–2444.

Muller, H.J. (1914). A gene for the fourth chromosome of Drosophila. J. Exp. Zool. *17*, 325–336.

Muller, H.J. (1925). Why polyploidy is rarer in animals than in plants. Am. Nat. *59*, 346–353.

Nagai, J., Yamato, K.T., Sakaida, M., Yoda, H., Fukuzawa, H., and Ohyama, K. (1999). Expressed sequence tags from immature female sexual organ of a liverwort, Marchantia polymorpha. DNA Res. *6*, 1–11.

Nishiyama, R., Yamato, K.T., Miura, K., Sakaida, M., Okada, S., Kono, K., Takahama, M., Sone, T., Takenaka, M., Fukuzawa, H., and Ohyama, K. (2000). Comparison of expressed sequence tags from male and female sexual organs of Marchantia polymorpha. DNA Res. *7*, 165–174.

Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., et al. (2003). Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution. Proc. Natl. Acad. Sci. USA *100*, 8007–8012.

Nishiyama, T., Wolf, P.G., Kugita, M., Sinclair, R.B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K., et al. (2004). Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol. Biol. Evol. *21*, 1813–1819.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature *497*, 579–584.

O'Donoghue, M.T., Chater, C., Wallace, S., Gray, J.E., Beerling, D.J., and Fleming, A.J. (2013). Genome-wide transcriptomic analysis of the sporophyte of the moss Physcomitrella patens. J. Exp. Bot. *64*, 3567–3581.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., et al. (1992). Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J. Mol. Biol. *223*, 1–7.

Ohtaka, K., Hori, K., Kanno, Y., Seo, M., and Ohta, H. (2017). Primitive Auxin Response without TIR1 and Aux/IAA in the Charophyte Alga Klebsormidium nitens. Plant Physiol. *174*, 1621–1632.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., et al. (1986). Chloroplast Gene Organization Deduced from Complete Sequence of Liverwort Marchantia polymorpha Chloroplast DNA. Nature *322*, 572–574.

Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., Kubo, M., and Hasebe, M. (2009). A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. Proc. Natl. Acad. Sci. USA *106*, 16321–16326.

Oostendorp, C. (1987). The Bryophytes of Paleozoic and the Mesozoic. Bryophyt. Bibl. *34*, 1–110.

Ortiz-Ramírez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijó, J.A., and Becker, J.D. (2016). A transcriptome atlas of Physcomitrella patens provides insights into the evolution and development of land plants. Mol. Plant *9*, 205–220.

Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., et al. (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA *104*, 7705–7710.

Panchy, N., Lehti-Shiu, M., and Shiu, S.H. (2016). Evolution of Gene Duplication in Plants. Plant Physiol. *171*, 2294–2316.

Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. Nature *424*, 194–197.

Pickett-Heaps, J.D. (1969). The evolution of the mitotic apparatus: an attempt at comparative ultrastructuralcytology in the dividing plant cells. Cytobios *1*, 257–280.

Proust, H., Hoffmann, B., Xie, X., Yoneyama, K., Schaefer, D.G., Yoneyama, K., Nogué, F., and Rameau, C. (2011). Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss Physcomitrella patens. Development *138*, 1531–1539.

Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA *103*, 15511–15516.

Rambaut, A. (2014). FigTree-v1.4.2. http://treebioedacuk/software/figtree/.

Renault, H., Alber, A., Horst, N.A., Basilio Lopes, A., Fich, E.A., Kriegshauser, L., Wiedemann, G., Ullmann, P., Herrgott, L., Erhardt, M., et al. (2017). A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nat. Commun. *8*, 14713.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64–69.

Rink, W. (1935). Zur Entwicklungsgeschichte, Physiologie und Genetik der Lebermoosgattungen Anthoceros und Aspiromitus. Flora *130*, 87–130.

Rizzon, C., Ponger, L., and Gaut, B.S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput. Biol. *2*, e115.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539–542.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., Malinovsky, F.G., Tör, M., de Vries, S., and Zipfel, C. (2011). The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. Plant Cell *23*, 2440–2455.

Rövekamp, M., Bowman, J.L., and Grossniklaus, U. (2016). Marchantia MpRKD Regulates the Gametophyte-Sporophyte Transition by Keeping Egg Cells Quiescent in the Absence of Fertilization. Curr. Biol. *26*, 1782–1789.

Rüdinger, M., Polsakiewicz, M., and Knoop, V. (2008). Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. Mol. Biol. Evol. *25*, 1405–1414.

Sakata, Y., Komatsu, K., and Takezawa, D. (2014). ABA as a Universal Plant Hormone. Prog. Bot. *75*, 57–96.

Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA. Genome Res. *10*, 516–522.

Sanderfoot, A. (2007). Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants. Plant Physiol. *144*, 6–17.

Sarris, P.F., Cevik, V., Dagdas, G., Jones, J.D.G., and Krasileva, K.V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biol. *14*, 8.

Schönherr, J., and Ziegler, H. (1975). Hydrophobic cuticular ledges prevent water entering the air pores of liverwort thalli. Planta *124*, 51–60.

Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107.

Sheard, L.B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T.R., Kobayashi, Y., Hsu, F.F., Sharon, M., Browse, J., et al. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. Nature *468*, 400–405.

Shimamura, M. (2016). Marchantia polymorpha: Taxonomy, Phylogeny and Morphology of a Model System. Plant Cell Physiol. *57*, 230–256.

Shiu, S.H., and Bleecker, A.B. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc. Natl. Acad. Sci. USA *98*, 10763–10768.

Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res. *41*, E344–E347.

Smit, A.F.A., and Hubley, R. (2008–2015). RepeatModeler Open-1.0. <http://www.repeatmasker.org>.

Smit, A.F.A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0. <http://www.repeatmasker.org>.

Smith, S.A., and Pease, J.B. (2017). Heterogeneous molecular processes among the causes of how sequence similarity scores can fail to recapitulate phylogeny. Brief. Bioinform. *18*, 451–457.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S. (2009). Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. *37*, 7002–7013.

Stumpe, M., Göbel, C., Faltin, B., Beike, A.K., Hause, B., Himmelsbach, K., Bode, J., Kramell, R., Wasternack, C., Frank, W., et al. (2010). The moss Physcomitrella patens contains cyclopentenones but no jasmonates: mutations in allene oxide cyclase lead to reduced fertility and altered sporophyte morphology. New Phytol. *188*, 740–749.

Suetsugu, N., Takemiya, A., Kong, S.G., Higa, T., Komatsu, A., Shimazaki, K., Kohchi, T., and Wada, M. (2016). RPT2/NCH1 subfamily of NPH3-like proteins is essential for the chloroplast accumulation response in land plants. Proc. Natl. Acad. Sci. USA *113*, 10424–10429.

Suire, C., Bouvier, F., Backhaus, R.A., Bégu, D., Bonneu, M., and Camara, B. (2000). Cellular localization of isoprenoid biosynthetic enzymes in Marchantia polymorpha. Uncovering a new role of oil bodies. Plant Physiol. *124*, 971–978.

Szövényi, P. (2016). The Genome of the Model Species Anthoceros agrestis. In Genomes and Evolution of Charophytes, Bryophytes, Lycophytes, and Ferns, S. Rensing, ed. (Academic Press), pp. 189–211.

Szövényi, P., Rensing, S.A., Lang, D., Wray, G.A., and Shaw, A.J. (2011). Generation-biased gene expression in a bryophyte model system. Mol. Biol. Evol. *28*, 803–812.

Takuno, S., Ran, J.H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. Nat Plants *2*, 15222.

Tan, X., Calderon-Villalobos, L.I.A., Sharon, M., Zheng, C., Robinson, C.V., Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature *446*, 640–645.

Tanaka, M., Esaki, T., Kenmoku, H., Koeduka, T., Kiyoyama, Y., Masujima, T., Asakawa, Y., and Matsui, K. (2016). Direct evidence of specific localization of sesquiterpenes and marchantin A in oil body cells of Marchantia polymorpha L. Phytochemistry *130*, 77–84.

Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. Nat. Rev. Genet. *12*, 692–702.

Tougane, K., Komatsu, K., Bhyan, S.B., Sakata, Y., Ishizaki, K., Yamato, K.T., Kohchi, T., and Takezawa, D. (2010). Evolutionarily conserved regulatory mechanisms of abscisic acid signaling in land plants: characterization of ABSCISIC ACID INSENSITIVE1-like type 2C protein phosphatase in the liverwort Marchantia polymorpha. Plant Physiol. *152*, 1529–1543.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Tsuzuki, M., Nishihama, R., Ishizaki, K., Kurihara, Y., Matsui, M., Bowman, J.L., Kohchi, T., Hamada, T., and Watanabe, Y. (2016). Profiling and Characterization of Small RNAs in the Liverwort, Marchantia polymorpha, Belonging to the First Diverged Land Plants. Plant Cell Physiol. *57*, 359–372.

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. *18*, 411–424.

Villarreal A, J.C., Crandall-Stotler, B.J., Hart, M.L., Long, D.G., and Forrest, L.L. (2016). Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytol. *209*, 1734–1746.

Wang, L., and Roossinck, M.J. (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. Plant Mol. Biol. *61*, 699–710.

Wang, C., Liu, Y., Li, S.S., and Han, G.Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. Plant Physiol. *167*, 872–886.

Wang, D., Zhang, Y., Zhang, S., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics *8*, 77–80.

Weirauch, M.T., and Hughes, T.R. (2011). A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. Subcell. Biochem. *52*, 25–73.

Wellman, C.H., Osterloff, P.L., and Mohiuddin, U. (2003). Fragments of the earliest land plants. Nature *425*, 282–285.

Wheeler, J.A. (2000). Molecular Phylogenetic Reconstructions of the Marchantioid Liverwort Radiation. Bryologist *103*, 314–333.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA *111*, E4859–E4868.

Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K.L., Stewart, N.R., Syrenne, R.D., Yang, X., Gao, P., et al. (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics *10* (*Suppl 11*), S3.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y., et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science *343*, 1505–1508.

Yamamoto, Y., Ohshika, J., Takahashi, T., Ishizaki, K., Kohchi, T., Matusuura, H., and Takahashi, K. (2015). Functional analysis of allene oxide cyclase, MpAOC, in the liverwort Marchantia polymorpha. Phytochemistry *116*, 48–56.

Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T., et al. (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. USA *104*, 6472–6477.

Yeh, R.F., Lim, L.P., and Burge, C.B. (2001). Computational inference of homologous gene structures in the human genome. Genome Res. *11*, 803–816.

Yue, J., Hu, X., Sun, H., Yang, Y., and Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. Nat. Commun. *3*, 1152.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D.G., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell *125*, 749–760.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Amborella trichopoda genome V1.0	(Albert et al., 2013)	http://amborella.huck.psu.edu/
Arabidopsis thaliana genome TAIR V10	Phytozome	http://genome.jgi.doe.gov
Picea abies genome V1.0	(Nystedt et al., 2013)	ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0
Selaginella moellendorffii genome V1.0	Phytozome; (Banks et al., 2011)	http://genome.jgi.doe.gov
Physcomitrella patens genome V3.0	Phytpzome; (Rensing et al., 2008)	http://genome.jgi.doe.gov
Spirogyra pratensis transcriptome	(Ju et al., 2015)	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSM01&search=GBSM01000000&display=scaffolds
Coleochaete orbicularis transcriptome	(Ju et al., 2015)	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSL01&search=GBSL01000000&display=scaffolds
Nitella mirabilis trancriptome	(Ju et al., 2015)	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBST01&search=GBST01000000&display=scaffolds
Klebsormidium nitens genome	(Hori et al., 2014; Ohtaka et al., 2017)	http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/kf_download.htm
Mesostigma viride transcriptome	(Ju et al., 2015)	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSK01&search=GBSK01000000&display=scaffolds
Chlamydomonas reinhardtii genome	Phytozome; (Merchant et al., 2007)	https://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV10
Ostreococcus tauri genome	(Derelle et al., 2006)	https://bioinformatics.psb.ugent.be/gdb/ostreococcus/
Ostreococcus lucimarinus genome	Phytozome; (Palenik et al., 2007)	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
charophyte and Chlorophyte transcriptomes	(Cooper and Delwiche, 2016)	https://figshare.com/articles/Green_algal_transcriptomes_for_phylogenetics_and_comparative_genomics/1604778
NCBI Conserved Domain Database	(Marchler-Bauer et al., 2015)	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
PROSITE	(Sigrist et al., 2013)	http://prosite.expasy.org
Carbohydrate-Active enZYmes Database	(Lombard et al., 2014)	http://www.cazy.org/
Experimental Models: Organisms/Strains		
Marchantia polymorpha subsp ruderalis Tak-1	this study	N/A
Marchantia polymorpha subsp ruderalis Tak-2	this study	N/A
Marchantia polymorpha subsp ruderalis Tak-1 BC4	this study	N/A
Marchantia polymorpha subsp ruderalis Kit-2	this study	N/A
Software and Algorithms		
Newbler (v2.6)	(Margulies et al., 2005)	N/A
Arachne v.20071016	(Jaffe et al., 2003)	https://github.com/cseed/arachne-pnr
Burrows-Wheeler Alignment tool (*BWA*)	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
(Sequence Alignment/Map) SAMtools	(Li et al., 2009)	http://samtools.sourceforge.net/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gepard	(Krumsiek et al., 2007)	http://mips.gsf.de/services/analysis/gepard
BLAT - the BLAST-like alignment tool	(Kent, 2002)	https://www.soe.ucsc.edu/?kent
MIRA ver. 3.2.0	(Chevreux et al., 1999)	https://sourceforge.net/projects/mira-assembler/
PERTRAN	(S. Shu, unpublished data)	N/A
Program to Assemble Spliced Alignments (PASA)	(Haas et al., 2003)	https://pasapipeline.github.io/
RepeatMasker	(Smit et al., 2013–2015)	http://www.repeatmasker.org/
RepeatModeler	(Smit and Hubley, 2008–2015)	http://www.repeatmasker.org/RepeatModeler.html
FGENESH+	(Salamov and Solovyev, 2000)	N/A
GenomeScan	(Yeh et al., 2001)	http://genes.mit.edu/genomescan/
BRAKER1	(Hoff et al., 2016)	http://bioinf.uni-greifswald.de/bioinf/braker/
CasFinder	N/A	http://arep.med.harvard.edu/CasFinder
STAR alignment software	(Dobin et al., 2013)	http://code.google.com/p/rna-star/
Cuffdiff	(Trapnell et al., 2013)	https://omictools.com/cuffdiff-tool
codonw 1.4.4 software	J. Peden	http://codonw.sourceforge.net
Spearman's rank correlation analysis tool version 1.1.23-r7	P. Wessa, Free Statistics Software, Office for Research Development and Education	https://www.wessa.net/
EMBOSS utility "getorf"	N/A	http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html
KOG	(Koonin et al., 2004)	https://www.ncbi.nlm.nih.gov/COG/new/shokog.cgi; ftp://ftp.ncbi.nih.gov/pub/COG/KOG
OrthoMCL	(Chen et al., 2006; Li et al., 2003)	http://orthomcl.org/orthomcl/
OMA	(Altenhoff et al., 2015)	http://omabrowser.org/oma/home/
MUSCLE v3.8.31	(Edgar, 2004)	http://www.ebi.ac.uk/Tools/msa/muscle/
KaKs_Calculator2.0	(Wang et al., 2010)	https://sourceforge.net/projects/kakscalculator2/
BLASTZ	(Schwartz et al., 2003)	genomewiki.ucsc.edu/index.php/Blastz
DAGchainer	(Haas et al., 2004)	http://dagchainer.sourceforge.net/
SynMap at CoGe website	(Lyons et al., 2008)	https://genomevolution.org/CoGe/SynMap.pl
Transposon Annotation using Small RNAs (TASR)	(El Baidouri et al., 2015)	https://tasr-pipeline.sourceforge.net/
LTRharvest	(Ellinghaus et al., 2008)	http://www.zbh.uni-hamburg.de/forschung/arbeitsgruppe-genominformatik/software/ltrharvest.html
LTRdigest	(Steinbiss et al., 2009)	http://www.zbh.uni-hamburg.de/forschung/arbeitsgruppe-genominformatik/software/ltrdigest.html
HMMER	N/A	http://hmmer.org/
Pfam	(Finn et al., 2016)	http://pfam.xfam.org/
Repeatmasker	(Smit et al., 2013–2015)	http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker
Repeatlandscape	Juan Caballero, RepeatMasker.org/ Institute for Sytems Biology, 2012	https://github.com/caballero/RepeatLandscape
SNAP v2.1.1	N/A	https://www.hiv.lanl.gov
Geneious ver. 7.1.9 package	N/A	https://www.geneious.com
Se-Alv2.0a11	(Rambaut, 2014)	http://tree.bio.ed.ac.uk/software/seal/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RAxML ver. 8.2.8	(Stamatakis, 2014)	https://github.com/stamatak/standard-RAxML
MrBayes ver. 3.2.6	(Ronquist et al., 2012)	http://mrbayes.sourceforge.net/download.php
Other		
Genome portal: *Marchantia polymorpha v3.1*	this study	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha_er
Genome portal: *Marchantia polymorpha v3.1* Includes raw RNaseq & DNA methylation data	this study	http://marchantia.info/
Transcriptome data	this study	Table S2 Mendeley https://doi.org/10.17632/zb7hwyj3hp.1
Marchantia nomenclature website	this study; (Bowman et al., 2016a)	http://marchantia.info/nomenclature/index.php/Main_Page
Phylogenetic tree analyses	this study	Mendeley https://doi.org/10.17632/zb7hwyj3hp.1 Tables S8-10

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact John L. Bowman (john.bowman@monash.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Several accessions of *M. polymorpha* subsp. ruderalis (Bischler-Causse, 1993; Bowman et al., 2016a; Shimamura, 2016) were utilized. The Y chromosome of the Takaragaike-1 (Tak-1) accession was previously sequenced (Yamato et al., 2007). The genome sequence presented here was derived from a single clonal female derived from backcross 4 between a male Tak-1 line and a female Takaragaike-2 (Tak-2) line (GenBank BioSample: SAMN02199054; Sample name: 33950; SRA: SRS441030). Genome sequence was also obtained from a female of the Kitashirakawa-2 (Kit-2) accession to facilitate identification of X chromosome scaffolds (GenBank BioSample: SAMN05920745; DOE Joint Genome Institute: Gp0033047). Sporeling transcriptome data were derived from a cross between the Cambridge-1 (Cam-1) and Cambridge-2 (Cam-2) accessions. Takaragaike and Kitashirakawa accessions were collected in Sakyo-ku, Kyoto, Japan, in which the International Conference Hall and Kyoto University (university botanical garden) are located, respectively. Cam accessions were collected in Cambridge, UK.

METHOD DETAILS

Sequencing Summary

Using a whole genome shotgun sequencing strategy, we sequenced *Marchantia polymorpha*, with DNA isolated from a single clonal female derived from backcross 4 between a male Tak-1 line, whose Y chromosome was previously sequenced (Yamato et al., 2007), and a female Tak-2 line. Sequencing reads were collected using 454, Illumina and Sanger sequencing platforms using standard sequencing protocols. The 454, Illumina, and Sanger reads were sequenced at the Department of Energy (DOE) Joint Genome Institute (JGI) in Walnut Creek, California and the HudsonAlpha Institute in Huntsville, Alabama. 454 reads were sequenced using the GS FLX+ platform, Illumina reads were sequenced using the Illumina MiSeq/HiSeq, and Sanger reads were sequenced using an ABI 3730XL capillary sequencer. Five linear 454 libraries (26.03x), one 2x250, 800bp insert Illumina fragment library (39.93x), one 2x150 5-kb insert Illumina mate pair library (13.38x), and three fosmid libraries (0.81/0.66x) were obtained for a total of approximately 80x coverage (see Table S11). Prior to assembly, all reads were screened for mitochondria, chloroplast, and phiX. Reads composed of > 95% simple sequence repeats were removed. Illumina reads < 75bp after trimming for adaptor and quality ($q < 20$) were removed. An additional deduplication step was performed on Illumina pairs that identifies and retains only one copy of each PCR duplicate.

Genome Assembly Process

The current release is the version 3.1 release which is a combination of the version 0.6 and the version 1.0 release. The process of assembly and combination to produce the version 3.1 release is detailed in this section. The version 0.6 release used a total of 26.03x linear 454 data combined with 0.81x Fosmid end data and assembled using Newbler (v2.6) (Margulies et al., 2005) with default

settings. This produced a raw assembly consisting of 4,618 scaffolds (8,623 contigs) totaling 201.7 Mb of sequence, with a scaffold N50 of 1.3 Mb, 223 scaffolds larger than 100 kb (180.4 Mb). Scaffolds were removed from the release if they were (a) repetitive; defined as scaffolds smaller than 50 kb consisting of > 95% 24mers that occurred 4 or more times in scaffolds larger than 50 kb, (b) contained only unanchored RNA-sequences, (c) were less than 1 kb in length, or (d) mitochondria/chloroplast contaminants. The total release includes 201.1 Mb of sequence assembled into 4,396 scaffolds (8,401 contigs) with a contig N50 of 62.8 kb and a gap content of 10.9%.

The version 1.0 release was assembled using a total of 53.31x Illumina data combined with the 0.66x Fosmid data and assembled using our modified version of Arachne v.20071016 (Jaffe et al., 2003) with parameters correct1_passes = 0 maxcliq1 = 700 BINGE_AND_PURGE = True max_bad_look = 1000. This produced a raw assembly consisting of 4,353 scaffolds (8,944 contigs) totaling 218.6 Mb of sequence, with a scaffold N50 of 1.5 Mb, 227 scaffolds larger than 100 kb (207.4 Mb). The screening procedure used with version 0.6 was applied to version 1.0. The total release includes 213.7 Mb of sequence assembled into 1,317 scaffolds (4,960 contigs) with a contig N50 of 262.1 kb and a gap content of 5.7%.

The version 2.0 release is a combination of the version 0.6 and version 1.0. The more complete version 1.0 assembly was used to fill gaps in the version 0.6 assembly. A total of 2,632 gaps were patched in version 0.6 resulting in a total of 8,942,510 bp added to the version 0.6 release. The patching increased the contig N50 from 62.8 kb to 256.9 kb. Post patching, a number of the version 0.6 scaffolds were rendered redundant. A total of 1,469 redundant scaffolds totaling 5.7 Mb of version 0.6 scaffolds were removed.

The version 1.0 assembly was masked using 24mers taken from version 0.6 and this revealed that there were a significant number of $\geq$ 1 kb contiguous sequences present in version 1.0 that were not present in the version 0.6 assembly. A total of 111 contigs (362 kb) were identified in version 1.0 and were combined with the original version 0.6 assembly to form the version 2.0 release. X chromosome (sex chromosome) related scaffolds were identified by aligning Kit-2 and Tak-1 reads to the version 2.0 release using bwa (Li and Durbin, 2009) and the per base depth was computed using samtools (Li et al., 2009). The Kit-2 and Tak-1 varieties were selected to be of different sexes and X chromosome related regions were identified as regions where the Tak-1 depth was ~0 while the Kit-2 depth was unchanged. A total of 7 scaffolds representing 4.2 MB of sequence were identified as X chromosome related. Homozygous SNPs and INDELs were corrected in the version 2.0 release using ~150x of illumina reads (2x250, 800 bp insert). Homozygous SNPs or INDELs that were within 50bp of one another were not corrected. A total of 35 homozygous SNPs was corrected out of a possible 546 and 5,585 indels were corrected out of a total 6,290. Finally, the version 3.1 release was formed by combining the version 2.0 release with the two available Y chromosome scaffolds [labeled Chr_Y_A and Chr_Y_B, (Yamato et al., 2007)]. The final version 3.1 assembly consists of 2,957 scaffolds (4,454 contigs) covering 225.8 Mb of the *Marchantia polymorpha* genome with a contig L50 of 265.9 kb and a scaffold L50 of 1.4 Mb. See Table S11 for the final version 3.1 scaffold and contigs statistics.

Assessment of Assembly Accuracy

A set of 19 PAC clones was sequenced in order to assess the accuracy of the assembly (Table S11). Minor variants were detected in the comparison of the PAC clones and the assembly. In 16 of the 19 PAC clones, the alignments were of high quality (< 0.03% bp error) with an overall error rate of 1 in 16,239 bp. An example of a high-quality PAC alignment is given in Table S11. The overall bp error rate (including marked gap bases) in all 19 PAC clones is 0.062% (1,287 discrepant bp out of 2,073,539). Table SM1c shows the individual PAC clones and their contribution to the overall error rate. Note that three PAC clones (4084730, 4084731, & 4084728) contribute 92% of the discrepant bases. This is due to one of the clones (4084728) terminating in a repetitive region (Table S11) and the discrepancies in the other clone are due to an insertion in the genome assembly (4084731) shown in Table S11.

Assessment of Completeness

Completeness of the euchromatic portion of the genome assembly was assessed using 1,682,726 454 ESTs (see below for their origin). The aim of this analysis is to obtain a measure of completeness of the assembly, rather than a comprehensive examination of gene space. The ESTs were aligned to the assembly using BLAT (Kent, 2002), with parameters: -t = dna –q = rna –extendThroughN. Alignments $\geq$ 90% base pair identity and $\geq$ 60% coverage were retained. The screened alignments indicate that 1,519,194 (95.93%) of the ESTs aligned to the assembly. Out of the remaining ESTs, 102,216 (6.07%) were library artifacts (chimera, or < 50% aligned at high identity), 3,081 (0.19%) sequences had $\geq$ 50 percent coverage with low identity, and 61,235 (3.87%) sequences were not found. The ESTs that were not found were checked against the NCBI non-redundant nucleotide repository (nr), and > 60% of them appear to be fungal plant pathogens and commensal prokaryotes.

As a second approach to assess completeness of v3.1, we compared this assembly to a preliminary assembly based on long reads (Pacbio). We first compared the length of the genome assembled by each of the approaches (Table S11). The length of the v3.1 (minus the Y chromosome of 6 Mb) is 219 Mb, while the female Pacbio assembly is 224 Mb. There are approximately 15 Mb of Ns in v3.1 (repetitive DNA that was not assembled into contigs correctly), while the Pacbio assembly lacks these. Conversely, the Pacbio assembly has about 13 Mb more 'repeat masked' bases (Table S11). Thus, the total length of the non-repetitive DNA portion of the two assemblies is very similar.

To assess the completeness of the Pacbio assembly, we asked whether it covers the 'gene space' present in the v3.1 assembly (Table S11). Thus, we mapped the annotated genes in v3.1 to the Pacbio assembly and found that of the 19,138 protein coding genes, all but 127 (99.3%) were in the Pacbio assembly; this indicates that the Pacbio assembly is also essentially complete. We then asked whether there is any 'gene space' in the Pacbio assembly not present in the v3.1 assembly. We mapped all transcriptome reads from

10 transcriptome experiments to the v3.1 (+/−Y chromosome) and Pacbio female assemblies. In every case v3.1 (w/o Y) had a slightly higher mapping rate than the Pacbio female. The Pacbio sequence failed to identify any additional functional genes on the X chromosome scaffolds. In summary, the v3.1 assembly accurately reflects the entire gene space in *M. polymorpha*.

Sequencing of Organellar Genomes

Nucleotide sequences of the organellar genomes of BC4 and Kit-2 were determined by assembling 454 and Illumina reads, respectively. Sequence discrepancy was observed only in poly A/T stretches of BC4 and Kit-2, suggesting the organellar genomes in the two accessions were conserved (Table S1). The sequences were mapped onto the published *Marchantia* organellar sequences, NC_001319.1 and NC_001660.1 for plastid and mitochondrial genomes, respectively. Collected reads were assembled by MIRA ver. 3.2.0 (Chevreux et al., 1999) run on the supercomputer of the Academic Center for Computing and Media Studies, Kyoto University, Japan). The sequences were finished by filling gaps with Sanger reads.

Annotation Methods

32,718 transcript assemblies were made from ~0.5B pairs of paired-end Illumina RNA-seq reads, 24,803 transcript assemblies from ~0.35B single-end Illumina RNA-seq reads (derived from *XVE:miR-E(z)* lines (Flores-Sandoval et al., 2016) and ectopic Mp*KNOX2* and Mp*BELL* expression lines (Tom Dierschke and J.L.B., unpublished data) to enrich for sporophyte expressed genes), 19,438 transcript assemblies from ~3M 454 reads (Table S1). All these transcript assemblies from RNA-seq reads were made using PERTRAN (S. Shu, unpublished data). 43,727 transcript assemblies were constructed using PASA (Haas et al., 2003) from 33,863 Sanger ESTs (Nagai et al., 1999; Nishiyama et al., 2000; Yamato et al., 2007) and all RNA-seq transcript assemblies above. Loci were determined by transcript assembly alignments and/or EXONERATE alignments of proteins from *Arabidopsis thaliana*, soybean, sorghum, rice, mimulus, grape, *Chlamydomonas reinhardtii*, *Sphagnum fallax*, *Physcomitrella patens* and *Klebsormidium nitens*, Swiss-Prot proteomes, and Pfam and Panther filtered ORFs from transcriptome assemblies of 4 Charophycean algae (*Nitella*, *Spirogyra*, *Coleochaete*, and *Mesostigma*) to repeat-soft-masked *Marchantia polymorpha* genome using RepeatMasker (Smit et al., 2013–2015) with up to 2-kb extension on both ends unless extending into another locus on the same strand. Repeat library was generated using RepeatModeler (Smit and Hubley, 2008–2015). Gene models were predicted by homology-based predictors, FGENESH+ (Salamov and Solovyev, 2000), FGENESH_EST (similar to FGENESH+, EST as splice site and intron input instead of protein/translated ORF), and GenomeScan (Yeh et al., 2001), and from AUGUSTUS via BRAKER1 (Hoff et al., 2016). The best-scored predictions for each locus were selected using multiple positive factors including EST and protein support, and one negative factor: overlap with repeats. The selected gene predictions were improved by PASA. Improvement includes adding UTRs, splicing correction, and adding alternative transcripts. PASA-improved gene model proteins were subject to protein homology analysis to above-mentioned proteomes to obtain C-score and protein coverage. C-score is a protein BLASTP score ratio to MBH (mutual best hit) BLASTP score and protein coverage is highest percentage of protein aligned to the best of homologs. PASA-improved transcripts were selected based on C-score, protein coverage, EST coverage, and its CDS overlapping with repeats. The transcripts were selected if its C-score is larger than or equal to 0.5 and protein coverage larger than or equal to 0.5, or it has EST coverage, but its CDS overlapping with repeats is less than 20%. For gene models, whose CDS overlaps with repeats for more than 20%, its C-score must be at least 0.9 and homology coverage at least 70% to be selected. The selected gene models were subjected to Pfam analysis and gene models whose protein was more than 30% in Pfam TE domains were removed. To define transcriptional units on the genome, we constructed full-length enriched cDNA libraries from thalli grown under different light conditions and sporeling. Both ends of 33,000 clones from each library were sequenced.

Sporeling RNA Enrichment Analysis

Crosses: *Marchantia polymorpha* Cam-1 & Cam-2 gametophytes were grown in soil and induced by far-red LED exposure for sexual organ development. Sperm was collected from Cam-1 antheridia and sprayed onto Cam-2 archegoniophores every 2-3 days for 1 week. After 3 weeks, sporangia began emerging and spore heads (archegoniophores bearing sporangia) were collected and dried with silica gel on 50mL Falcon tubes for 3 days. Spore heads were stored at −80°C. *Spore germination and axenic growth*: Twenty frozen spore heads were crushed and resuspended with 1mL of sterile water per spore head. Resuspended spores were filtered through a 40uM cell strainer and spun down at 13.000 RPM for 1 minute. Spores were sterilized by adding 1mL of Milton's sterilizing solution (1 Milton mini sterilizing tablet in 25mL of sterile water, Milton Pharmaceutical UK Company, active ingredient: Sodium dichloroisocyanurate CAS: 2893-78-9: 19.5% w w-1) and incubated at RT for 20 minutes in a rotating shaker at 100RPM. Spores were then spun down and washed once with 1mL of water. Finally, they were resuspended in 100uL per spore head and sown onto 1/2 Gamborg's B5 plates with 1% agar in decreasing amounts in respect to the expected mass obtained through germination (1/3 of total spores were kept for time 0h, 1/4 of total spores for time 24h, 1/5 for time 48h, 1/8 for time 72h and 1/10 for time 96h). Plates were sealed with micropore tape and placed in the culture chamber on a 16:8 day/night cycle at 21°C, with 60 μmol photons m−2 s−1 of illumination. *Sample collection and sequencing*: Spores were collected from plates at the same hour of the day, every 24 hours for 4 days using sterile water with an L-spreader, and placed on 1.5mL Eppendorf tubes. Excess liquid was removed and tubes stored at −80°C. Three biological replicates were performed for each time point. Total RNA was extracted using QIAGEN RNeasy mini kit and RNA concentration was measured with a Qubit fluorimeter using the RNA BR assay kit. RNA integrity was assessed with a Bioanalyzer 2100 machine using the Agilent RNA 6000 Nano kit and samples were sent to BGI for further processing.

Library preparation was performed with the Illumina TruSeq RNA library preparation kit v.2 and 100bp paired-end sequencing was carried out on an Illumina HiSeq 2000 machine, with samples multiplexed on 2 lanes. *RNA sequencing:* RNA sequencing produced > 24M clean paired-end (PE) reads per time point on average. Reads were mapped to the *Marchantia polymorpha* v3.1 genome using the STAR alignment software (Dobin et al., 2013) and mapped reads were processed through Cuffdiff (Trapnell et al., 2013). Sporeling FPKM normalized read counts were then loaded onto the FPKM matrix as sporeling datasets and enrichment analysis for sporelings was carried out by obtaining the ratio between the maximum value of any sporeling dataset over the value of any other tissue (see Table M1 at Mendeley https://doi.org/10.17632/zb7hwyj3hp.1). Genes were defined as enriched in sporeling tissue if the ratio was > 10. Also, genes having infinity ratios (where all other tissues had 0 FPKM values) were assigned as enriched if the maximum expression value of the sporeling datasets was > 0.3 FPKM, as to avoid extremely low counts from being assigned as enriched. This approach produced 366 genes enriched in sporeling tissue, which can be found in the supplementary enriched genes table.

Codon Analysis

METHODOLOGY: Analyses of indices of codon usage across the transcriptome dataset were performed using codonw 1.4.4 software (J. Peden, http://codonw.sourceforge.net). Correlation analysis was carried out using a Spearman's rank correlation analysis tool (P. Wessa, Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, https://www.wessa.net/).

Identification of Upstream ORFs

The full transcript sequences, including splice variants, for *A. trichopoda*, *A. thaliana*, *M. polymorpha*, *P. patens* and *S. moellendorffii* were obtained from Phytozome v11.0, TAIR 10 and ENSEMBLE databases. We predicted all ORFs in the forward strand for each sequence with EMBOSS utility "getorf," detecting the longest ORF as the coding gene and filtering the 5′ upstream uORFs if the size in aminoacids were > 9 using custom Perl scripts (available upon request), all uORFs must have an AUG and a stop codon within the 5′UTR. Information for initial and end positions, uORF size and uORF distance to ORF was recorded in text tables (see Table S1). The minimum size was set to 10 amino acids, including the start and stop codon, this consideration was taken due to the experimental evidence from functional uORFs. We considered all splicing variants to perform our analysis, all uORFs described here must have an AUG and a stop codon within the 5′UTR, in other organisms there a great number of uORFs that overlap with the main ORF, in plants most of the uORFs described and for those with functional experimental evidence are located on the 5′UTR. As a positive control we were able to identify 100% of the Conserved Peptide uORFs reported in *A. thaliana* previously.

Ortholog Analysis

To examine conservation of orthologs and expansion and loss of protein families, we performed ortholog analysis using three resources, KOG (Koonin et al., 2004), OrthoMCL (Chen et al., 2006; Li et al., 2003), and OMA (Altenhoff et al., 2015). They are designed to cluster homologs according to similarity scores obtained by an all-against-all search (BLAST for KOG and OrthoMCL; Smith-Waterman for OMA). Although it becomes more difficult to identify genuine orthologs as sequences to be examined become more distant from their most recent common ancestor (Smith and Pease, 2017), they provide a practical way to compare orthologs and protein families among multiple organisms. Of the 19,138 genes, 13,175 were assigned to at least one group by at least one of the three resources (Table S3). The numbers and members of groups varied among the three resources and we found several groups whose members are not consistent with those identified by rigorous phylogenetic analysis (e.g., in OMA: phytochrome and several families of transcription factors). Thus, only results from OrthoMCL and KOG are presented in Table S3. Due to the incongruences in the different ortholog analyses, perhaps due to the phylogenetic distances involved and sparse taxon sampling (Smith and Pease, 2017), we chose to perform more rigorous phylogenetic analyses for specific gene families of interest (see main text and subsequent Supplemental Information).

Whole and Segmental Genome Duplication Analysis

The transcript with the longest ORF among transcriptional variants was collected from each gene. Paralog pairs were selected as reciprocal best hits by BLAST. The amino acid sequences of the pairs were aligned by MUSCLE (v3.8.31; Edgar, 2004), and their Ks and 4dTv values were calculated by the KaKs_Calculator2.0 [Table S4, (Wang et al., 2010)]. Dot plot figures generated using BLASTZ (Schwartz et al., 2003) and DAGchainer (Haas et al., 2004) included in SynMap at CoGe website (Lyons et al., 2008).

Identification of Repetitive DNA

We performed *de novo* annotation of repeats through a homology based approach by building a library of consensus sequences with RepeatModeler (that combines both RepeatScout and RECON) (Smit and Hubley, 2008–2015) and subsequently employing the consensus sequences as probes to screen for similar sequences with RepeatMasker (Smit et al., 2013–2015). In addition, we employed Transposon Annotation using Small RNAs (TASR) (El Baidouri et al., 2015) by taking advantage of the different *M. polymorpha* small RNA libraries publically available. Repeat elements were identified using the RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html) pipeline that combines and automates four different programs RECON, RepeatScout, RepeatMasker (http://www.repeatmasker.org/) and Tandem Repeats Finder (TRF) to build, refine, and classify consensus models of putative interspersed repeats (Lerat, 2010). Long Terminal Retrotransposons (LTRs) were identified using the LTRharvest (Ellinghaus et al., 2008) LTRdigest (Steinbiss et al., 2009) pipeline. LTR retrotransposons were subjected to computational identification and

manual curation to identify LTRs containing particular arrays of protein domains characteristics of retrotransposon proteins using HMMER (http://hmmer.org/) and Pfam (http://pfam.xfam.org/) database for domain search. Assignment of superfamilies was done based on the array of the Integrase (INT), and Reverse Transcriptase (RT) domains. Filtered LTRs were classified into Copia (GAG AP INT RT RH) and Gypsy (GAG AP RT RH INT) families. This way we identified 264 full-length LTRs. Consensus libraries derived from both pipelines were employed to determine the repetitive content in the *M. polymorpha* genome using Repeatmasker and Repeatlandscape (https://github.com/caballero/RepeatLandscape; http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html).

Sex Chromosome Analysis

X chromosome scaffolds were identified by aligning Kitashirakawa-2 (Kit-2) and Tak-1 reads to the genome assembly. The Kit-2 (female) and Tak-1 (male) varieties were selected to be of different sexes and X chromosome regions were identified as regions where the Tak-1 depth was ~0 while the Kit-2 depth was unchanged. A total of 9 scaffolds, two of which (17, 18) are large, representing 4.37 Mb of sequence were identified as X chromosome related (Table S6). Consistently, all available X-linked markers (Fujisawa et al., 2001; Yamato et al., 2007) were mapped onto the scaffolds.

Since the sex chromosomes are diverged not only in sequence but also in function, the expression profiles of the sex chromosomal genes were examined. RNA-seq data from vegetative and reproductive tissues from each sex (see Table S6) were mapped to the genomic sequence by tophat ver. 2.1.0, and gene expression profiles were estimated by cuffdiff ver. 2.2.1 (Trapnell et al., 2012). A given gene is "reproductive," when its expression level (FPKM/RPKM) is larger than 0.1 in reproductive tissues, antheridiophore or antheridium in Tak-1 (male) and archegoniophore, archegonium or sporophyte in Tak-2 (female), and its vegetative to reproductive expression ratio is larger than 10. Synonymous versus non-synonymous rates are analayzed with SNAP v2.1.1 (https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html).

Phylogenetic Methods

Protein and/or transcript sequences were collected using the *Marchantia* genome portal site (http://marchantia.info/genome/index.php/Main_Page) based on the sequence similarity, or from the GenBank databases at NCBI. In some case, domains were evaluated using NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) or PROSITE (http://prosite.expasy.org).

Multiple sequence alignments were performed using the MUSCLE program (Edgar, 2004) contained in the Geneious ver. 7.1.9 package (https://www.geneious.com) with the default settings. Alternatively, amino acid sequences were manually aligned using Se-Alv2.0a11. After removing alignment gaps manually or using the Strip Alignment Columns routine in the Geneious package, phylogenetic analyses were performed using RAxML ver. 8.2.8 (Stamatakis, 2014) or MrBayes ver. 3.2.6 (Ronquist et al., 2012) programs. Alignments used to generate the phylogenetic trees are available upon request.

Predicted protein sequences obtained from the sequenced genomes of land plants, charophytes, and Chlorophytes and from the transcriptomes of charophytes listed below. In some cases for which greater resolution with the charophycean algal grade was desired, additional algal sequences were interrogated (Cooper and Delwiche, 2016). Finally, sequences in GenBank from additional algal or land plant taxa, as detailed in figure legends, were utilized for the examination of select gene families (Table S11).

Arabidopsis thaliana

TAIR Ver. 10
proteins
https://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV10
Athaliana > annotation > Athaliana_167_TAIR10.protein_primaryTranscriptOnly.fa.gz
Ref: http://www.arabidopsis.org

Amborella trichopoda

Ver. 1.0
proteins
http://amborella.huck.psu.edu/
Data > Filtered peptide sequences
Ref: (Albert et al., 2013)

Picea abies

Ver. 1.0
transcripts; proteins
ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/FASTA/Z4006_Transcripts/454.newbler.mRNA.AllContigs.fna>
ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/FASTA/GenePrediction/
Pabies1.0-HC-pep.faa.gz; Pabies1.0-MC-pep.faa.gz; Pabies1.0-LC-pep.faa.gz
Ref: (Nystedt et al., 2013)

Selaginella moellendorffii
Ver. 1.0
proteins
https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
Smoellendorffii > annotation > Smoellendorffii_91_v1.0.protein_primaryTranscriptOnly.fa.gz
Ref: (Banks et al., 2011).

Physcomitrella patens
Ver. 3.0 proteins
https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
Ppatens > annotation > Ppatens_251_v3.0.protein_primaryTranscriptOnly.fa.gz
Ref: (Rensing et al., 2008)

Spirogyra pratensis transcripts
https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSM01&search=GBSM01000000&display=scaffolds
GBSM01.1.fsa_nt.gz
Ref: (Ju et al., 2015)

Coleochaete orbicularis transcripts
https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSL01&search=GBSL01000000&display=scaffolds
GBSL01.1.fsa_nt.gz
Ref: (Ju et al., 2015)

Nitella mirabilis transcripts
https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBST01&search=GBST01000000&display=scaffolds
GBST01.1.fsa_nt.gz
Ref: (Ju et al., 2015)

Klebsormidium nitens*
Ver. 1.0
proteins
http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/kf_download.htm
Ref: (Hori et al., 2014)
*Originally described as *K. flaccidum* NIES-2285 for which the draft genome sequence was produced (Hori et al., 2014), and was taxonomically reclassified as *K. nitens* NIES-2285 (Ohtaka et al., 2017). The name change has bee corrceted in the text, but the phylogenetic trees may still contain names reflecting the older nomenclature of Hori et al. (2014).

Mesostigma viride
transcripts
https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GBSK01&search=GBSK01000000&display=scaffolds
GBSK01.1.fsa_nt.gz
Ref: (Ju et al., 2015)

Chlamydomonas reinhardtii
Ver. 5.5
proteins
https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
Creinhardtii > annotation > Creinhardtii_281_v5.5.protein_primaryTranscriptOnly.fa.gz
Ref: (Merchant et al., 2007)

Ostreococcus tauri
proteins
https://bioinformatics.psb.ugent.be/gdb/ostreococcus/
Ostta_prot_LATEST.tfa.gz
Ref:(Derelle et al., 2006)

Ostreococcus lucimarinus
Ver. 2.0
proteins
https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
Olucimarinus > annotation > Olucimarinus_231_v2.0.protein_primaryTranscriptOnly.fa.gz
Ref: (Palenik et al., 2007)

To standardize the appearance of the phylogenetic trees presented, the primary taxa from which sequences were obtained have been color-coded. All algal genes are blue, with charophyte sequences being readily distinguishable from Chlorophyte sequences, by a dark blue versus a light blue appearance, respectively. All *M. polymorpha* sequences are in bold black and are denoted by their Mapoly numbers. *P. patens* genes are in green, with vascular plant genes in orange/brown/purple/red.

Arabidopsis thaliana angiosperm At red
Amborella trichopoda angiosperm Am, AmTr purple
Picea abies gymnosperm Pa brown
Selaginella moellendorffii lycophyte Sm, Smoe orange
Physcomitrella patens moss Pp, Phpat green
Marchantia polymorpha liverwort Mp, Mapoly **black**
Spirogyra pratensis charophyte Sp, Spipr dark blue
Coleochaete orbicularis charophyte Co, Color dark blue
Nitella mirabilis charophyte Nm, Nitmi dark blue
Klebsormidium nitens charophyte Kf, Kfl dark blue
Mesostigma viridae charophyte Mv, Mesvi dark blue
Chlamydomonas rienhardtii Chlorophyte Cr, Cre light blue
Ostreococcus lucimarinus Chlorophyte Ol, Olu light blue
Ostreococcus tauri Chlorophyte Ot light blue

Sequences from taxa different than those listed above are color-coded with their respective phylogenetic placement, e.g., angiosperms, red/purple; gymnosperms, brown; ferns, light brown; lycophytes, orange; liverworts, black. If additional taxa are utilized, their color codes are listed in the respective figure legends.

Note that due the limited number of species of Chlorophytes, charophyte algae, and Embryophytes used in this study, the origins of gene families are dated in a manner analogous to inferred dates in the fossil record. The proposed origins represent the latest possible dates, with earlier origins possible; additional genome and transcriptome sequences will refine the estimated times of origins reported here. Some phylogenetic trees are presented in Figure 4 and Tables S8-10; additional tree data are available from Mendeley https://doi.org/10.17632/zb7hwyj3hp.1.

ABA Analyses

Transformation of A. thaliana pyr1pyl1pyl2pyl4 quadruple mutant: A cDNA of Mp*PYL1* fused in frame to the N terminus of sGFP cDNA was introduced into pCambia1300, adjacent to the cauliflower mosaic virus 35S promoter, and used for transformation by *Agrobacterium* infiltration of the inflorescence. Transgenic seedlings were selected on agar plates containing 30 mg/L hygromycin and 100 mg/L cefotaxime. *Immunoblot analysis:* Leaves of transgenic T2 plants were used for immunoblot analysis for confirmation of MpPYL1-GFP accumulation. Protein extract was electrophoresed, blotted onto PVDF membrane, and reacted with anti-GFP antibody (#598, MBL, Nagoya, Japan). Positive signals were detected by horseradish peroxidase-conjugated secondary antibody (#458, MBL, Nagoya, Japan) and the chemi-luminescence reagent (Chemi-lumi One, Nacalai Tesque, Japan). *Germination tests:* Seeds of transgenic T2 plants were sown on plates of B5 medium with or without 1.5 μM ABA. The sown seeds were stratified at 4°C for one week, and then grown at 23°C under continuous light for seven days.

QUANTIFICATION AND STATISTICAL ANALYSES

Genome sequencing, accuracy and completeness of assembly, and annotation, along with sporeling and sex chromosome gene transcriptome quantification can be found in the relevant sections of the Methods Details.

DATA AND SOFTWARE AVAILABILITY

Genomic resources are available from web sites of both JGI (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha_er) and the *Marchantia* user community (http://marchantia.info/). Both sites provide similar services, such as a genome browser, BLAST and keyword searches, but there are some features unique to each. The site of JGI, for example, is a part of Phytozome and thus suitable for comparative genomics, and bulk data can be downloaded. On the other hand, the site by the user community is more oriented to supporting researchers who are using *M. polymorpha*. Its genome browser provides separate

tracks for the RNA-seq data (see Table S2 and Mendeley https://doi.org/10.17632/zb7hwyj3hp.1) and methylation, and a platform for community annotation. Unified gene nomenclature is facilitated through a community website, (http://marchantia.info/nomenclature/index.php/Main_Page), with nomenclatural rules outlined (Bowman et al., 2016a). Additional phylogenetic analyses are available at Mendeley, https://doi.org/10.17632/zb7hwyj3hp.1.

Raw sequence reads are available at https://www.ncbi.nlm.nih.gov/sra (SRX874572-SRX874573, SRX555320-SRX555475, SRX301553-SRX301560, SRX114614-SRX114615, SRX030759-SRX030787, SRX2268331-SRX2268345)

Additional phylogenetic analyses can be found at Mendeley, https://doi.org/10.17632/zb7hwyj3hp.1.

Supplemental Figures

A

Class	Taxa	Sequence	AP2 domain	B3 domain	ARF domain (MD)	PB1 domain	TAIR BLASTP
Mesostigmatophyceae	*Mesostigma viride*	comp26332_c2_seq1					AT5G60450.1 | Symbols: ARF4 |
		comp18232_c0_seq1					AT1G13260.1 | Symbols: RAV1, EDF4 |
		comp40667_c0_seq1					AT3G61970.1 | Symbols: NGA2 |
Klebsormidiophyceae	*Entransia*	comp17988_c0_seq1					AT2G28350.1 | Symbols: ARF10
		comp25035_c0_seq1					AT1G01030.1 | Symbols: NGA3 |
	Klebsormidium flaccidum	kfl00094_0070					AT1G13260.1 | Symbols: RAV1,
Charophyceae	*Nitella mirabilis*	comp76386_c0_seq1					AT2G28350.1 | Symbols: ARF10
		comp99450_c2_seq1					AT5G06250.2 | Symbols: | AP2
Coleochaetophyceae	*Chaetosphaeridium globosum*	comp34242_c0_seq1					AT2G28350.1 | Symbols: ARF10
		comp21460_c0_seq1					AT5G06250.2 | Symbols: | AP2
	Coleochaete orbicularis	comp30027_c2_seq1					AT1G19850.1 | Symbols: MP, AR
		comp15661_c0_seq1					AT2G28350.1 | Symbols: ARF10
		comp27966_c0_seq1					AT1G13260.1 | Symbols: RAV1,
Zygnematophyceae	*Penium margaritaceum*	comp25890_c0_seq1					AT5G37020.1 | Symbols: ARF8,
	Spirogyra pratensis	comp16086_c0_seq1					AT4G30080.1 | Symbols: ARF16
		comp7997_c0_seq1					AT1G13260.1 | Symbols: RAV1,
	Spirogyra sp Au1	comp11493_c2_seq1					AT4G30080.1 | Symbols: ARF16
		comp11502_c0_seq1					AT1G13260.1 | Symbols: RAV1,
	Mougeotia scalaris	comp10166_c0_seq1					AT1G30330.1 | Symbols: ARF6 |
		comp10411_c0_seq1					AT1G30330.1 | Symbols: ARF6 |
		comp14239_c0_seq1					AT2G46870.1 | Symbols: NGA1 |
Marchantiopsida	*Marchantia polymorhpa*	Mapoly0019s0045.1					AT1G30330.2 | Symbols: ARF6 |
		Mapoly0011s0167.1					AT5G62000.3 | Symbols: ARF2,
		Mapoly0043s0098.1					AT2G28350.1 | Symbols: ARF10
		Mapoly0072s0102.1					AT1G13260.1 | Symbols: RAV1, EDF4 |

B

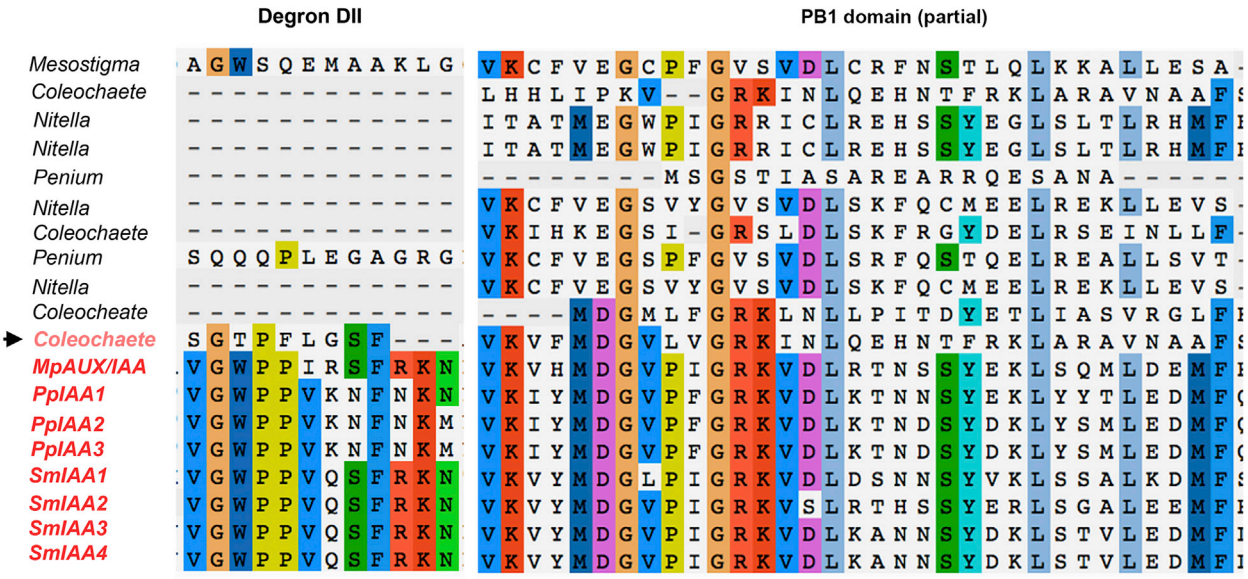


Figure S1. Evolution of ARF and AUX/IAA families, Related to Figure 5

(A) Three ARF clades (A, B, C) exist in land plants, with at least two (A/B, C) existing in a charophyte ancestor (see also B3 phylogentic tree at Mendeley, https://doi.org/10.17632/zb7hwyj3hp.1). ARFs with B3, ARF-MID and PB1 domains evolved within, or prior to the emergence of the charophycean algae.

(B) Searches for charophyte AUX/IAA-like sequences identified genes with IAA-like PB1 domains, but that were not associated with B3 domains. However, none of the charophyte IAA-like genes has either conserved domains I or II (required for TPL or TIR1 interaction). A single *Coleochaete* gene has a proto-domain II with some but not all the canonical amino acids necessary for TIR interaction, suggesting that *IAA* genes evolved from a B3-PB1 gene via loss of a B3 domain and, likely subsequently acquiring DI and DII and the capacity to interact with TPL and TIR1 in the ancestral land plant.

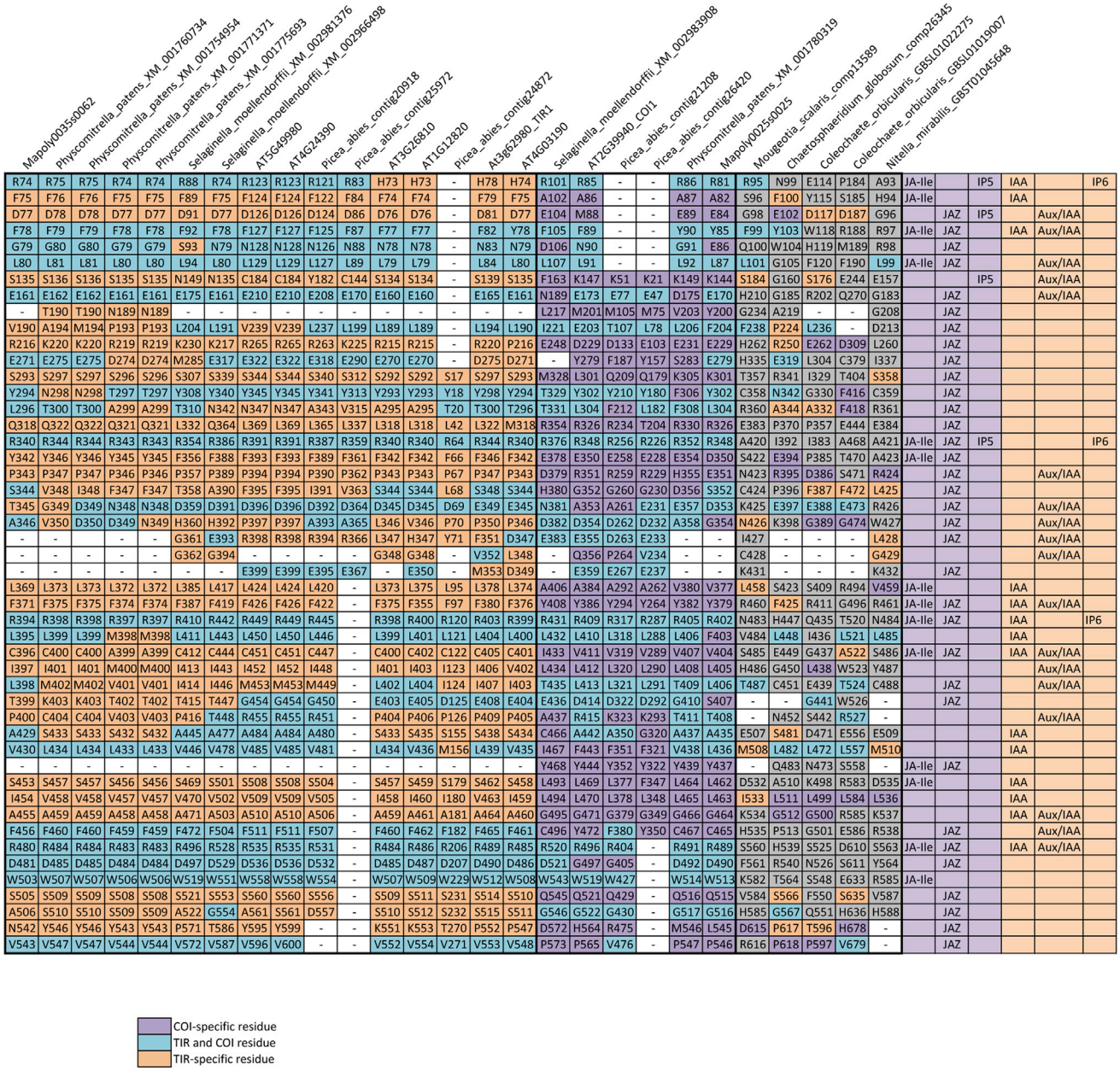


Figure S2. Conservation of AtTIR1/AFBs and COI1 Residues Involved in Signaling Interactions, Related to Figure 5

Conservation of AtTIR1/AFBs and COI1 residues involved in interactions with auxin, Aux/IAA, IP6 (inositol hexakisphosphate), JA-Ile, JAZ or IP5 (inositol pentakisphosphate) in TIR1, and COI1 orthologs and related algal F-box proteins. Sequences were aligned with MUSCLE. The right part of the table shows the role of the amino acids in the interaction with IAA, Aux/IAA, IP6 in TIR1 or with JA-Ile, JAZ and IP5 (inositol pentakis-phosphate) in COI1. Numbers indicate positions within each protein. Color code is indicated in the key.

Cell

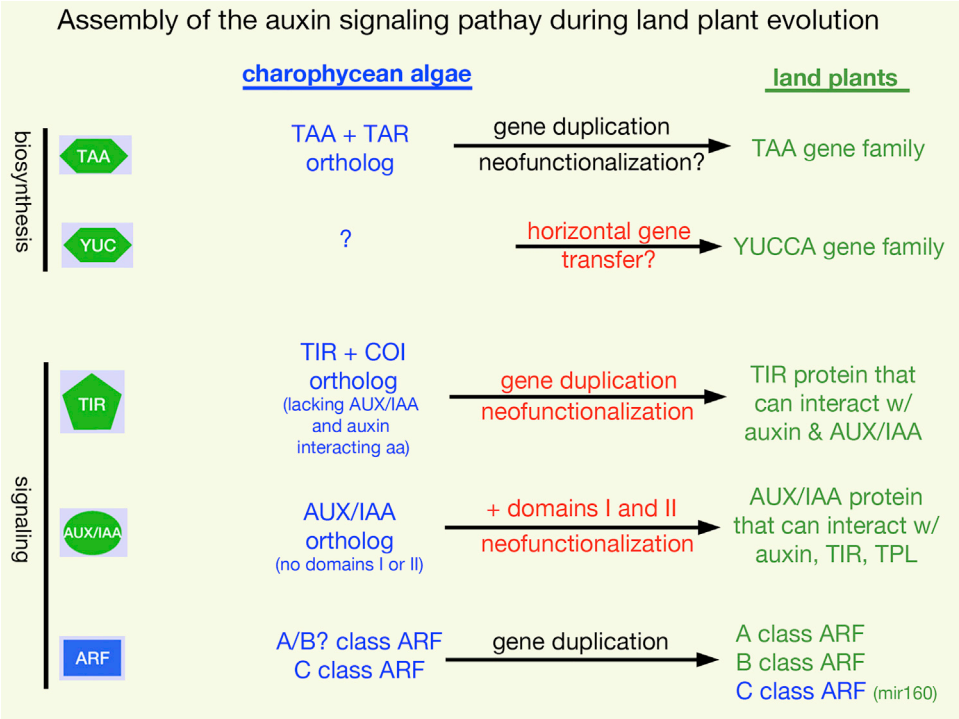


Figure S3. Summary of the Proposed Origins of Components of Auxin Biosynthesis and Signaling Components, Related to Figure 5

The assemby of the auxin transcriptional signaling pathway in an ancestral land plant exemplifies the origins of new genes and functions that evolved in the land plant lineage. With respect to auxin biosynthesis, the land plant TAA gene family and has its origin in a gene duplication, possibly followed by neofunctionalization producing paralogs with new activities. In contrast, convincing YUCCA orthologs were not identifed in available charophycean algal databases — the origin of land plant YUC genes may have been via horizontal gene transfer from a microbe. The auxin receptor, TIR1, has its origin in a gene duplication followed by neofunctionalization into TIR1 and COI proteins that possess the ability to interact with auxin and AUX/IAA proteins (TIR1) or a jasmonate-related ligand and JAZ proteins (COI). Neofunctionalization of an AUX/IAA precursor via addition of domain II enabled the evolution of an AUX/IAA-TIR complex that interacts in the presence of auxin, thus assembling the auxin transcriptional signaling system. Acquisition of domain I by AUX/IAA genes enabled the repressive capability of AUX/IAA proteins. ARF transcription factors predated the evolution of the auxin-signaling system, suggesting they could retain ancestral functions independent from auxin. Thus, the auxin biosynthesis and signaling components were assembled via modification (neofunctionalization) of both pre-existing and newly evolved genes.

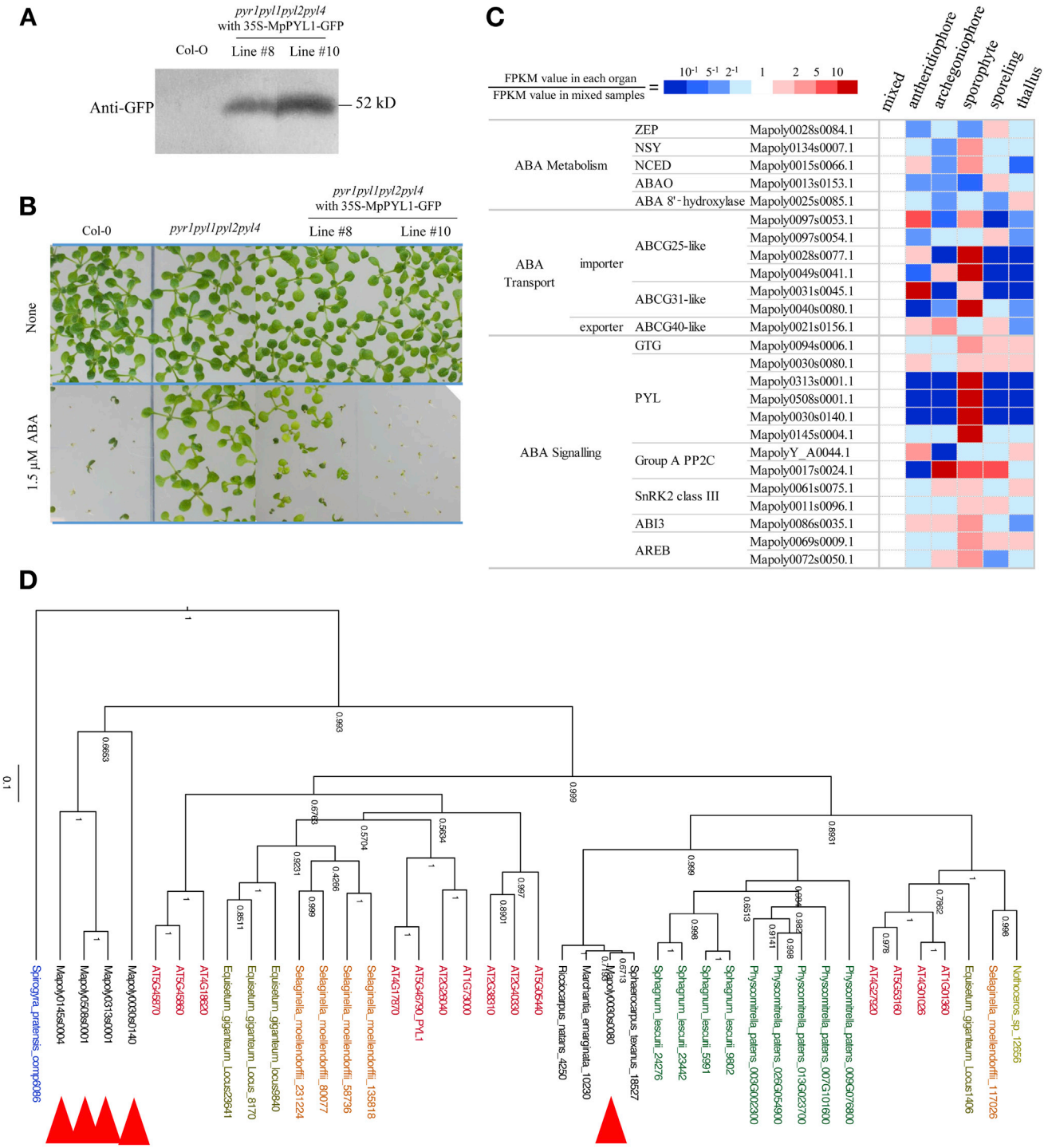


Figure S4. ABA Signaling Pathway, Related to Figure 5

(A and B) Complementation by Mp*PYL1* (Mapoly0030s0080) of mutants of ABA receptors in *A. thaliana*. The *35S:MpPYL1-GFP* construct was introduced into the quadruple mutant (*pyr1pyl1pyl2pyl4*) of ABA receptors by Agrobacterium-mediated transformation. Leaves of the generated transgenic plants were used for immunoblot analysis using anti-GFP antibody to determine the protein expression levels (A). Two lines (#8 and #10) expressing different amounts of MpPYL1-GFP were used for germination tests on the agar medium containing 1.5 μM ABA (B). Photo was taken after 7 days of culture following stratification of seeds.

(C) Comparison of transcriptional profiles for ABA-related genes in various organs of *M. polymorpha.* Comparison of transcriptional profiles for ABA-related genes. Color chart shows relative FPKM values. The mixed organ sample is assigned a value of 1 for every gene, while all other samples are presented with values relative to the mixed sample. Non-detect FPKM values were calculated as 10^{-6}.

(legend continued on next page)

(D) Phylogram of PYR/PYL/RCAR proteins. Amino acid sequences were aligned and BEAST was used to generate the tree. Numbers on the branches indicate posterior probability values. While there are five *M. polymorpha* genes, Mapoly0030s0080 (Mp*PYL1*) is phylognetically more similar to angiosperm proteins shown to be ABA receptors. The other *M. polymorpha* genes are phylogenetically more diverse that the *Arabidopsis* proteins, and experiments are required to determine if they are bona fide ABA receptors. The most similar gene in the charophycean algal transcriptomes is Spirogyra_pratensis_comp6086, which differs from the consensus at several sites in the ABA binding pocket and the gate and latch (see Table S8); again whether this is a bona fide ABA receptor must be determined experimentally.